Máster en Tecnologías y Sistemas de Comunicaciones
Depto. de Señales, Sistemas y Radiocomunicaciones, UPM

Curso "Métodos analíticos y análisis de señales"

Módulo A
Métodos algebraicos para el estudio de señales y sistemas

9 de diciembre de 2015

José Ignacio Ronda Prieto
jir@gti.ssr.upm.es
www.gti.ssr.upm.es/~jir

Grupo de Tratamiento de Imágenes
Depto. de Señales, Sistemas y Radiocomunicaciones
ETSI Telecomunicación, Universidad Politécnica de Madrid

Índice general

Preámbulo

Este documento es la primera parte de los apuntes del curso de doctorado "Métodos analíticos y análisis de señal"del Máster Universitario en Tecnologías y Sistemas de Comunicaciones de la ETSIT-UPM.

El objetivo del curso es reforzar los recursos matemáticos de los ingenieros de telecomunicación para facilitar la realización de la tesis doctoral.

En esta primera parte se intenta facilitar el uso del álgebra lineal como herramienta en esta rama de la ingeniería. Esta parte del curso se divide en tres partes:

- En los primeros temas, básicamente de repaso y nivelación, se aprovecha para establecer conexiones entre conceptos de álgebra lineal y de teoría de la señal.
- A continuación se estudian el análisis de componentes principales, la descomposición en valores singulares y varias versiones del problema de mínimos cuadrados, temas que probablemente constituyen las herramientas fundamentales para abordar problemas de análisis de señales en términos de subespacios y distancias euclídeas. Los fundamentos proporcionados permiten abordar de forma sencilla otros problemas como el análisis discriminante lineal y el escalado multidimensional.
- En los últimos temas se estudian las cuestiones fundamentales relativas a la implementación de algoritmos matriciales, como son ciertas factorizaciones matriciales y los conceptos de condicionamiento y estabilidad.

Capítulo 1

Introducción al álgebra matricial

1.1. Espacios vectoriales. Bases

En este capítulo incluimos algunas cuestiones básicas de álgebra matricial que resultarán útiles en los desarrollos del curso.

Un *espacio vectorial* V sobre un cuerpo $\mathbf{K}$ es un conjunto cuyos elementos denominaremos *vectores* dotado de dos operaciones: la suma de vectores, que representaremos con el símbolo de suma habitual, y el producto de vector por un elemento del cuerpo *(escalar),* que representaremos con un punto o sin operador. Las operaciones deben verificar las siguientes propiedades (representando en negrita los vectores):
(1) $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$ (asociativa)
(2) $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$ (conmutativa)
(3) $\mathbf{u}+\mathbf{0}=\mathbf{u}$ para todo $\mathbf{u}$ (existencia de elemento neutro)
(4) Para todo $\mathbf{u}$ existe $-\mathbf{u}$ tal que $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$ (existencia de elemento simétrico)
Las propiedades anteriores (1), (3) y (4) son las que caracterizan a un *grupo*, y la (2) la que lo hace un grupo *conmutativo.*
(5) $\alpha(\beta\mathbf{u})=(\alpha\beta)\mathbf{u}$
(6) $1\mathbf{u}=\mathbf{u}$
(7) $\alpha(\mathbf{u}+\mathbf{v})=\alpha\mathbf{u}+\alpha\mathbf{v}$ (distributiva).

Ejemplos de espacios vectoriales son $\mathbf{R}^n$ (sobre $\mathbf{R}$) y $\mathbf{C}^n$ (sobre $\mathbf{C}$) definiendo la suma y el producto por escalar de la forma natural. Otros ejemplos: El conjunto de las funciones de variable real, el de las funciones continuas de variable real, el de las secuencias. En la práctica, a la hora de comprobar si un conjunto es un espacio vectorial lo primero que hay que verificar es que las operaciones estén bien definidas, es decir, que cuando multiplicamos por un escalar o cuando sumamos dos elementos no nos salimos del conjunto.

Un *subespacio vectorial* o *variedad lineal* es un subconjunto de un subespacio vectorial que a su vez es subespacio vectorial.

Un *sistema de generadores* es un conjunto de vectores tal que cualquier vector del espacio se puede poner como *combinación lineal* de los elementos del conjunto (suma (finita) de elementos del conjunto multiplicados por escalares). Un *sistema libre* de vectores (o de vectores *linealmente independientes)* es un conjunto en el que ningún elemento se puede poner como combinación lineal de los demás, o, equivalentemente, en el que la única forma de conseguir el $\mathbf{0}$ como combinación lineal es tomar todos los escalares igual a cero.

Una *base* es un conjunto ordenado que es a la vez sistema de generadores y sistema libre. Tiene la propiedad de que todo vector tiene una representación única como combinación lineal de los vectores de la base. Esta representación se denomina *coordenadas* del vector. Se puede demostrar que si un espacio vectorial tiene una base finita, entonces todas sus bases son finitas y tienen el mismo número de elementos. Llamamos a este número *dimensión* del espacio.

Espacios de dimensión finita son, además de los $\mathbf{K}^n$, el conjunto de las secuencias periódicas de periodo N o el de las señales de tiempo continuo periódicas de periodo T formadas por frecuencias menores que W.

Cuando un espacio no es dimensión finita, se dice que es de *dimensión infinita.* Es fácil encontrar ejemplos de espacios de dimensión infinita en los espacios de señal.

En $\mathbf{K}^n$ se define la *base canónica* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, en la que $\mathbf{e}_i$ es el vector con todos los elementos iguales a cero excepto el i-ésimo, que vale uno.

1.2. Aplicaciones lineales

1.2.1. Aplicaciones lineales y matrices

Una *aplicación lineal* u *homomorfismo* entre dos espacios vectoriales U y V es una aplicación f de U en V con la propiedad

$$f(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha f(\mathbf{u}) + \beta f(\mathbf{v}).$$

A partir de esta definición es inmediato que f está determinada por las imágenes de los vectores de una base $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ de U:

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{u}_i \Rightarrow f(\mathbf{x}) = \sum_{i=1}^{n} x_i f(\mathbf{u}_i).$$

Se comprueba fácilmente que la composición de aplicaciones lineales es otra aplicación lineal y que la inversa de una aplicación lineal, si existe (es decir, si la aplicación es biyectiva), también es una aplicación lineal.

Si en un espacio vectorial U sobre $\mathbf{K}$, de dimensión n, fijamos una base, la función que asigna a cada vector sus coordenadas respecto de la base es una aplicación lineal biyectiva de U en $\mathbf{K}^n$ que nos permite en cierto modo identificar U con K^n. Sin embargo no conviene olvidar que la identificación es un tanto arbitraria, pues viene dada por la base, aunque en algunos casos pueda haber bases que resulten "naturales" y hagan que la identificación sea (o parezca) más natural. El ejemplo trivial es la base canónica de $\mathbf{K}^n$.

Usando bases en los espacios de partida y de llegada, toda aplicación lineal entre espacios vectoriales de dimensiones n y m se puede ver como una aplicación lineal de $\mathbf{K}^n$ en $\mathbf{K}^m$, aunque su determinación concreta depende (perdón por la insistencia) de las bases utilizadas.

Si f es una aplicación lineal de $\mathbf{K}^n$ en $\mathbf{K}^m$, definimos la *matriz* $m \times n$ asociada como una tabla de m filas y n columnas cuyas columnas son las imágenes de los vectores de la base canónica de $\mathbf{K}^n$.

Definimos el producto de una matriz A, $m \times n$, por un vector de $\mathbf{K}^n$, como el vector de $\mathbf{K}^m$ dado por

$$\underbrace{\begin{pmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{pmatrix}}_{A} \underbrace{\begin{pmatrix}x_1 \\ \vdots \\ x_n\end{pmatrix}}_{\mathbf{x}} = x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n.$$

Obsérvese que si A es la matriz de la aplicación lineal $f : \mathbf{K}^n \to \mathbf{K}^m$, tenemos

$$f(\mathbf{x}) = A\mathbf{x},$$

lo que justifica la definición de la matriz y del producto de matriz por vector.

Podemos visualizar el producto de una matriz por un vector como un nuevo vector generado como combinación lineal de las columnas de la matriz con coeficientes dados por el vector.

Si U tiene dimensión n y V tiene dimensión m, la aplicación lineal $f : U \to V$ tiene, como hemos dicho, una aplicación asociada definida $\mathbf{K}^n$ en $\mathbf{K}^m$, que dependerá de las bases elegidas. Por tanto podemos hablar de la *matriz de la aplicación* f respecto de estas bases.

Un caso importante de aplicación lineal de $\mathbf{K}^n$ en $\mathbf{K}^n$ es el dado por un *cambio de base* en un espacio vectorial. Si B y B' son dos bases de U, la aplicación que asocia a las coordenadas respecto de B, $\mathbf{x}_B \in \mathbf{K}^n$, de un vector $\mathbf{x} \in U$ sus coordenadas respecto de

B', $\mathbf{x}_{B'} \in \mathbf{K}^n$, es una aplicación lineal de $\mathbf{K}^n$ en $\mathbf{K}^n$. De acuerdo con la definición de matriz asociada a una aplicación lineal de $\mathbf{K}^n$ en $\mathbf{K}^m$, las columnas de la matriz de la aplicación del cambio de base serán las coordenadas respecto de B' de los vectores que en la base B tienen por coordenadas los vectores de la base canónica, es decir, de los vectores de la base B.

Utilizaremos las siguientes definiciones y notación: Una matriz *diagonal* es la que tiene todos los elementos igual a cero excepto los de la diagonal principal, que son los que tienen los dos índices iguales). I es la *matriz identidad* (matriz cuadrada con unos en la diagonal principal y ceros en el resto). Si A es una matriz, $A^\top$ es su *traspuesta* (la que obtenemos poniendo las columnas como filas), $\bar{A}$ su conjugada y A^* su traspuesta conjugada. Una matriz es *cuadrada* si tiene tantas filas como columnas, *simétrica* si coincide con su traspuesta y *hermítica* si coincide con su transpuesta conjugada. Una matriz *triangular superior (inferior)* es la que tiene sólo ceros debajo (encima) de la diagonal principal.

1.2.2. Producto de matrices

Consideramos las aplicaciones lineales

$$\mathbf{K}^n \xrightarrow{f} \mathbf{K}^m \xrightarrow{g} \mathbf{K}^p.$$

Si $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$ es la matriz de f y $B = \begin{pmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_m \end{pmatrix}$ la de g, definimos la *matriz producto* BA como la de la aplicación compuesta $g \circ f$. La columna k de BA será

$$g(f(\mathbf{e}_k)) = g(\mathbf{a}_k) = B\mathbf{a}_k.$$

Por tanto podemos visualizar el producto BA como una nueva matriz cuya columnas son combinaciones lineales de las de B mediante coeficientes dados por una columna de A:

$$AB = A \underbrace{\begin{pmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_r \end{pmatrix}}_{B} = \begin{pmatrix} A\mathbf{b}_1 & \dots & A\mathbf{b}_r \end{pmatrix}. \tag{1.1}$$

Y en el caso de que la segunda matriz sea una matriz columna (vector),

$$\underbrace{\begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\mathbf{x}} = x_1\mathbf{a}_1 + \ldots + x_n\mathbf{a}_n, \tag{1.2}$$

Igualmente útiles son las "versiones transpuestas" de estas fórmulas. La primera nos dice que el producto de una matriz fila por una matriz es un vector fila combinación lineal de las filas de la matriz con pesos indicados por los coeficientes del vector.

$$\begin{pmatrix} y_1 & \dots & y_m \end{pmatrix} \begin{pmatrix} \mathbf{b}_1^* \\ \vdots \\ \mathbf{b}_m^* \end{pmatrix} = \left(y_1 \mathbf{b}_1^* + \dots + y_n \mathbf{b}_n^* \right). \tag{1.3}$$

La segunda nos da el producto de dos matrices en términos de las filas de la primera:

$$\begin{pmatrix} \mathbf{a}_1^* \\ \vdots \\ \mathbf{a}_r^* \end{pmatrix} B = \begin{pmatrix} \mathbf{a}_1^* B \\ \vdots \\ \mathbf{a}_r^* B \end{pmatrix}. \tag{1.4}$$

El producto de dos matrices se puede calcular mediante la fórmula siguiente. Si $A = (a_{ij})$, $B = (b_{ij})$ y $AB = (c_{ij})$,

$$c_{ij} = \sum_k a_{ik} b_{kj}.$$

De la asociatividad de la composición de aplicaciones se desprende la asociatividad del producto de matrices.

Una propiedad inmediata es $(AB)^\top = B^\top A^\top$. Otra propiedad importante es que el producto se puede calcular *por bloques*, es decir, que si A y B se desglosan en submatrices A_{ij} y B_{ij} formadas seleccionando ciertas filas y columnas de la matriz global, siempre que los productos entre submatrices estén definidos, el producto AB se puede calcular aplicando la fórmula del producto a los bloques. Por ejemplo, si dos matrices 4×4 se subdividen en matrices 2×2 A_{ij} y B_{ij}, tenemos

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}.$$

1.2.3. Rango y núcleo de una matriz. Matriz inversa

La *imagen (range)* $\operatorname{im} A$ de una matriz A es el subespacio generado por sus columnas. El *núcleo (nullspace, kernel)* $\ker A$ de la matriz es el conjunto de vectores que tienen imagen cero.

El *rango (rank)* de una matriz es la dimensión de su imagen. Coincide con la dimensión del subespacio generado por sus filas (es decir, el rango de una matriz coincide con el de su traspuesta). El rango más la dimensión del núcleo es igual al número de columnas

(estas afirmaciones serán consecuencia inmediata del teorema de descomposición en valores singulares).

Una matriz $m \times n$ es de *rango máximo* si tiene el máximo rango posible dadas sus dimensiones, es decir, el mínimo de m y n.

Si una matriz cuadrada A, $m \times m$, es de rango máximo, entonces la dimensión de su imagen es m, es decir, es todo el espacio. En particular podremos expresar cada vector de la base canónica como combinación lineal de las columnas de A. De aquí se desprende que existe una matriz B tal que $AB = I$. B es la *matriz inversa* de A, que se nota A^{-1}, y que también verifica $A^{-1}A = I$. De hecho las condiciones $AB = I$ y $BA = I$ resultan ser equivalentes.

Una matriz cuadrada con inversa se llama *regular*, y en caso contrario se dice que es *singular*.

Se pueden demostrar las propiedades siguientes:
(1) $(AB)^{-1} = B^{-1}A^{-1}$,
(2) $(A^{-1})^\top = (A^\top)^{-1}$,
(3) $(A^{-1})^* = (A^*)^{-1}$.
Notaremos estas última matrices, respectivamente, como $A^{-\top}$ y A^{-*}.

Ejercicios

1.1. (a) Si $\mathbf{p}$ y $\mathbf{q}$ son vectores columna de cuatro componentes linealmente independientes, calcular el rango de la matriz $\mathbf{p}\mathbf{q}^\top - \mathbf{q}\mathbf{p}^\top$.
(b) Escribir esta matriz como un producto de matrices.

1.2. Utilizando la interpretación del producto de matrices, demostrar que el producto de dos matrices triangulares superiores es otra matriz triangular superior.

1.3. Dado el vector $\mathbf{v} \in \mathbf{R}^3$, obtener la matriz $M_\mathbf{v}$ tal que $M_\mathbf{v}\mathbf{x} = \mathbf{v} \times \mathbf{x}$ para cualquier vector $\mathbf{x}$. Indicar el la imagen y el núcleo de la matriz.

1.4. (a) Se dice que dos subespacios E y F de V son *complementarios* cuando el menor subespacio que contiene a ambos es V y su intersección es $\{0\}$. Demostrar que si E y F son complementarios todo vector de V se puede expresar de forma única como suma de un vector de E y otro de F.
(b) Una aplicación lineal P de V en V es una *proyección* cuando verifica $P^2 = P$.
Demostrar que el núcleo y la imagen de una proyección se cortan en $\{0\}$. Indicación: Tomar un $\mathbf{y} = P\mathbf{x}$ que esté en el núcleo y la imagen y demostrar que es cero
(c) Demostrar que cada vector se puede escribir de forma única como suma de un vector del núcleo de P y un vector de su imagen. Indicación: Escribir $\mathbf{x}$ como $\mathbf{x} = P\mathbf{x} + (\mathbf{x} - P\mathbf{x})$.
(d) Demostrar que si E y F son subespacios de V que se cortan en $\{0\}$ y expanden V, existe una única proyección que tiene a E y F, respectivamente, como núcleo y rango. (Indicación: Comprobar que la proyección restringida a su imagen es igual a la identidad.) Esta proyección se dice que es la *proyección*

sobre F paralela a E.
(d) Demostrar que si P es una proyección, también lo es $I - P$. ¿Qué relación existe entre los núcleos y los rangos de estas proyecciones?

1.3. Formas bilineales, sesquilineales y cuadráticas

Una *forma bilineal* ψ es una aplicación de V^2 en $\mathbf{K}$ lineal en ambas variables. Se dice que es *simétrica* si $\psi(\mathbf{u}, \mathbf{v}) = \psi(\mathbf{v}, \mathbf{u})$.

En el caso de espacios vectoriales sobre $\mathbf{C}$ hay que distinguir las formas bilineales simétricas de las *formas sesquilineales simétricas conjugadas*, que verifican:
(1) $\psi(\alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w}) = \alpha\psi(\mathbf{u}, \mathbf{w}) + \beta\psi(\mathbf{v}, \mathbf{w})$,
(2) $\psi(\mathbf{u}, \mathbf{v}) = \overline{\psi(\mathbf{v}, \mathbf{u})}$,
y, como consecuencia,
(3) $\psi(\mathbf{u}, \alpha\mathbf{v} + \beta\mathbf{w}) = \bar{\alpha}\psi(\mathbf{u}, \mathbf{v}) + \bar{\beta}\psi(\mathbf{v}, \mathbf{w})$.

Es fácil comprobar que una forma bilineal ψ está determinada por las imágenes de los pares de vectores de una base y que una forma bilineal en $\mathbf{K}^n$ se puede expresar como

$$\psi(\mathbf{u}, \mathbf{v}) = \mathbf{v}^\top M \mathbf{u},$$

donde M es simétrica si lo es ψ (y viceversa). De hecho, si $M = (m_{ij})$, $m_{ij} = \psi(\mathbf{e}_i, \mathbf{e}_j)$.

Una forma bilineal definida sobre otro espacio se puede expresar de la misma forma tomando una base. En ese caso los coeficientes de M son las imágenes de los pares de vectores de la base.

Análogamente, una forma sesquilineal simétrica conjugada definida sobre $\mathbf{C}^n$ está determinada por las mismas imágenes y se puede escribir como

$$\psi(\mathbf{u}, \mathbf{v}) = \mathbf{v}^* M \mathbf{u},\ M = (m_{ij}),\ m_{ij} = \psi(\mathbf{e}_i, \mathbf{e}_j).$$

Una *forma cuadrática* es una aplicación f de V en $\mathbf{K}$ definida a partir de una forma bilineal simétrica o de una forma sesquilineal simétrica conjugada ψ como

$$f(\mathbf{u}) = \psi(\mathbf{u}, \mathbf{u}).$$

Obsérvese que en el caso de que f derive de una forma sesquilineal simétrica conjugada, sólo puede tomar valores reales, puesto que la simetría conjugada implica $\psi(\mathbf{u}, \mathbf{u}) = \overline{\psi(\mathbf{u}, \mathbf{u})}$.

Respecto de una base, f está, obviamente, dada en el primer caso por una expresión de la forma

$$f(\mathbf{u}) = \mathbf{u}^\top M \mathbf{u},$$

con M simétrica, y en el segundo por una de la forma

$$f(\mathbf{u}) = \mathbf{u}^* M \mathbf{u},$$

con M hermítica. Una forma bilineal simétrica (o sesquilineal simétrica conjugada) está unívocamente definida por su forma cuadrática asociada (puesto que ésta determina la matriz M).

Las formas cuadráticas con valores en $\mathbf{R}$ se clasifican en *definida negativa, semidefinida negativa, nula, semidefinida positiva, definida positiva o indefinida* según tomen, para vectores no nulos, respectivamente, valores siempre negativos, negativos o nulos, nulos, positivos o nulos, siempre nulos o positivos y negativos.

1.4. Determinantes

Una *forma multilineal alternada* es una aplicación de V^r $(r \leq n)$ en $\mathbf{K}$ que es lineal en cada una de sus variables *(multilineal)* y cuyo valor cambia de signo si se intercambian los valores de dos variables *(alternada)*. Una forma multilineal alternada definida sobre V^n (es decir, que se come n vectores y produce un escalar) está unívocamente determinada por el valor que asigna a una base cualquiera de V.

El *determinante* de una tupla de n vectores de $\mathbf{K}^n$ es la forma multilineal alternada que asocia la unidad a la base canónica.

El *determinante* de una matriz cuadrada A, que se nota como $|A|$ o como $\det(A)$, es el determinante de sus vectores columna. Las principales propiedades del determinante son:
(1) $|AB| = |A||B|$
(2) $|A^\top| = |A|$
(3) A regular $\Leftrightarrow |A| \neq 0$.

1.5. Autovalores y autovectores

Una aplicación lineal de un espacio vectorial en sí mismo se denomina *endomorfismo*. Los *autovectores* de un endomorfismo f son vectores no nulos con la propiedad

$$f(\mathbf{v}) = \lambda \mathbf{v}.$$

Se dice entonces que el escalar λ es el *autovalor* asociado al autovector.

Si A es la matriz del endomorfismo de $\mathbf{K}^n$, un autovector $\mathbf{v}$ con autovalor λ estará en el núcleo de $A - \lambda I$. Por tanto $A - \lambda I$ es singular, luego su determinante es nulo. Por

tanto los autovalores de A son raíces del polinomio $|A-\lambda I|$, que se denomina *polinomio característico*.

Autovectores con distintos autovalores son linealmente independientes.

Cada raíz simple tiene asociado un autovector, que está definido salvo constante multiplicativa. Una raíz de multiplicidad n tiene asociado un espacio vectorial de dimensión $m \leq n$. Se dice que n es la *multiplicidad algebraica* del autovalor y que m es su *multiplicidad geométrica*.

Dos propiedades que relacionan los autovalores de una matriz con funciones de sus coeficientes son las siguientes:
(1) El producto de los autovalores es igual al determinante de la matriz.
(2) La suma de los autovalores es igual a la *traza* de la matriz (suma de los elementos de la diagonal principal).
Estas propiedades son consecuencia de las relaciones de Cardano aplicadas a los coeficientes del polinomio característico.

1.6. Buscando las matrices más sencillas

Dado que las matrices de las aplicaciones lineales y las formas cuadráticas dependen tanto de la aplicación o de la forma en sí como de las bases que tomemos, se plantea la cuestión de si podemos tomar bases respecto de las cuales las matrices tengan la forma más sencilla posible, que es la diagonal. Consideraremos tres casos por separado.

1.6.1. Aplicaciones lineales entre espacios distintos

Veamos en primer lugar cómo afectan los cambios de base a la matriz de una aplicación lineal.

Si la aplicación $f : U \to V$ tiene matriz A respecto de ciertas bases iniciales, y luego cambiamos estas bases de forma que los cambios de base vienen dados en U y en V, respectivamente, por las ecuaciones $\mathbf{x}' = C\mathbf{x}$, $\mathbf{y}' = D\mathbf{y}$, podemos obtener fácilmente la matriz de f respecto de las nuevas bases. Notando por $\mathbf{y}'$ las coordenadas respecto de la nueva base de la imagen del vector con nuevas coordenadas $\mathbf{x}'$, tenemos

$$\mathbf{y}' = D\mathbf{y} = DA\mathbf{x} = DAC^{-1}\mathbf{x}'$$

luego la matriz respecto de las nuevas bases es $A' = DAC^{-1}$.

Más adelante veremos que siempre podemos encontrar bases con matrices de cambio que hagan A' diagonal. Será una consecuencia inmediata del teorema de descomposición en valores singulares.

1.6.2. Diagonalización

Si f es un endomorfismo sólo podemos jugar con un cambio de base, por lo que como resultado de la discusión anterior tendremos $A' = CAC^{-1}$. Dos matrices relacionadas como lo están en este caso A y A' se dice que son *semejantes*.

Al tener menos grados de libertad para transformar A en A' que en el caso en que los espacios inicial y final son distintos, no se puede aplicar el resultado anterior. De hecho, existen endomorfismos para los que A' no se puede hacer diagonal. Se dice que una matriz es *diagonalizable* si existe un cambio de base que la hace diagonal, es decir, si es semejante a una matriz diagonal. Entonces las columnas de M son autovalores de A y los elementos correspondientes de D los autovalores asociados. Obsérvese que una matriz es diagonalizable si y sólo si existe una base de autovectores del endomorfismo asociado.

Por tanto una matriz es diagonalizable si y sólo si las multiplicidades geométricas de los autovalores coinciden con las algebraicas.

Si A se diagonaliza con una matriz V unitaria se dice que A es *unitariamente diagonalizable*. Se demuestra que las matrices unitariamente diagonalizables coinciden con las matrices *normales*, que son las que verifican $AA^* = A^*A$.

En el caso general la forma más sencilla que podemos obtener en estos casos es la *forma canónica de Jordan*, que en el caso complejo es una matriz diagonal con algunos unos encima de la diagonal principal. La obtención de la forma canónica de Jordan require aritmética exacta, como la que utilizamos cuando hacemos cálculos simbólicos, puesto que las perturbaciones aleatorias derivadas de la imprecisión numérica convierten con seguridad una matriz no diagonalizable en otra diagonalizable. No obstante, la forma canónica de Jordan tiene interés práctico en el estudio de fenómenos en los que la estructura del problema dé lugar necesariamente a matrices no diagonalizables.

1.6.3. Formas cuadráticas

Ahora consideramos una forma cuadrática asociada a una forma sesquilineal simétrica conjugada. El cambio de base dado por $\mathbf{u} = C\mathbf{u}'$ da lugar a un cambio en la matriz de la forma cuadrática que podemos obtener fácilmente:

$$\mathbf{u}^* M \mathbf{u} = \mathbf{u}'^* C^* M C \mathbf{u}' \Rightarrow M' = C^* M C.$$

Como veremos, siempre se puede encontrar una transformación de este tipo que hace M' diagonal.

Ejercicios

1.5. Si una matriz cuadrada se puede escribir como $A = VDV^{-1}$, donde D es diagonal, entonces las columnas de V son autovectores de A y los elementos correspondientes de D los autovalores asociados.

1.6. Se denomina *factorización de Schur* de una matriz cuadrada a una factorización de la forma

$$A = QTQ^*,$$

donde Q es unitaria y T es triangular superior.
(a) Comprobar que si $\mathbf{u}$ es un autovector unitario de A y U es una matriz unitaria con $\mathbf{u}$ como primera columna, el producto U^*AU es de la forma $\begin{pmatrix} \lambda & \mathbf{b}^* \\ \mathbf{0} & C \end{pmatrix}$. Comprobar que si $C = VT'V^*$, con V unitaria y T' triangular superior, y $Q = U\begin{pmatrix} 1 & \mathbf{0}^* \\ \mathbf{0} & V \end{pmatrix}$, entonces Q^*AQ es triangular superior.
(b) Demostrar que toda matriz cuadrada admite una factorización de Schur. Indicación: Operar por inducción en la dimensión de la matriz y utilizar el apartado anterior.

1.7. (a) Demostrar que una matriz normal triangular superior es necesariamente diagonal. Indicación: Operar por inducción.
(b) Demostrar que si una matriz es normal, la matriz triangular que aparece en su descomposición de Schur es normal.
(c) Demostrar que una matriz es unitariamente diagonalizable si y sólo si es normal.

1.8. Demostrar que la matriz $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ no es diagonalizable.

1.9. Las *matrices circulantes* son las que representan convoluciones circulares, y son de la forma

$$A = \begin{pmatrix} c_1 & c_2 & \dots & c_m \\ c_m & c_1 & \dots & c_{m-1} \\ \vdots & \vdots & & \vdots \\ c_2 & c_3 & \dots & c_1 \end{pmatrix}.$$

Demostrar que sus autovectores son

$$\mathbf{v}_k = \begin{pmatrix} 1 & \omega^{(k-1)} & \omega^{(k-1)2} & \dots & \omega^{(k-1)(m-1)} \end{pmatrix}^T,\ \omega = e^{j2\pi/m},\ k = 1, \dots, m,$$

y concluir que son unitariamente diagonalizables. ¿Qué relación existe entre la DFT y la diagonalización de una matriz circulante?

1.7. Espacios con producto escalar

1.7.1. Definiciones y propiedades básicas

En este apartado repasamos la estructura matemática que nos permite, entre otras cosas, decir si un vector es grande o pequeño, y si dos vectores están lejos o cerca (es decir, la estructura que hace posible Barrio Sésamo).

Veamos primero algunas estructuras más elementales relacionadas.

Un *espacio métrico* es un conjunto dotado de una *distancia*, que es una función d que asigna un valor real a cada par de elementos del conjunto, con las propiedades
(1) $d(x,y) = d(y,x)$ (simetría)
(2) $d(x,z) \leq d(x,y) + d(y,z)$ (desigualdad triangular)
(3) $d(x,y) = 0 \Leftrightarrow x = y$.

Una aplicación de un espacio vectorial V con valores en $\mathbf{R}^+$ es una *norma* si verifica las siguientes propiedades:
(1) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
(2) $\|\alpha\mathbf{u}\| = |\alpha|\|\mathbf{v}\|$
(3) $\|\mathbf{v}\| = 0 \Leftrightarrow \mathbf{v} = \mathbf{0}$.

Un espacio vectorial dotado de una norma se denomina *espacio normado*. La norma hace que el espacio vectorial sea también un espacio métrico.

Un *producto escalar* es una aplicación de $V \times V \to \mathbf{C}$ que verifica
(1) $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$ (simetría conjugada).
(2) $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{u}, \mathbf{w} \rangle + \beta \langle \mathbf{v}, \mathbf{w} \rangle$.
(3) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ y $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \Leftrightarrow \mathbf{u} = 0$.
De (1) y de (2) se deduce
$\langle \mathbf{u}, \alpha\mathbf{v} + \beta\mathbf{w} \rangle = \bar{\alpha} \langle \mathbf{u}, \mathbf{v} \rangle + \bar{\beta} \langle \mathbf{u}, \mathbf{w} \rangle$.
Estas propiedades se pueden resumir diciendo que un producto escalar es una forma sesquilineal simétrica conjugada definida positiva.

Otras propiedades derivadas de las anteriores son
$\mathbf{u} \perp \mathbf{v} \Rightarrow \|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ *(teorema de Pitágoras).*
$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2\,(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2)$ *(identidad del paralelogramo).*

Un espacio con producto escalar es un caso particular de espacio normado, con la norma $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$.

Recíprocamente, una norma deriva de un producto escalar si y sólo si verifica la igualdad del paralelogramo. En ese caso el producto escalar se puede definir a partir de la norma mediante la *identidad de polarización:*
$\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{4}(\|\mathbf{u} + \mathbf{v}\|^2 - \|\mathbf{u} - \mathbf{v}\|^2 - i(\|\mathbf{u} + i\mathbf{v}\|^2 - \|\mathbf{u} - i\mathbf{v}\|^2))$.

En $\mathbf{C}^n$ se define el *producto escalar canónico* como

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^* \mathbf{u}.$$

1.7.2. Sistemas ortonormales y matrices unitarias

Se dice que dos vectores son *ortogonales* si su producto escalar es cero, y ello se representa como $\mathbf{u} \perp \mathbf{v}$.

Un vector es *unitario* si su norma es uno. Un conjunto de vectores unitarios ortogonales dos a dos se denomina *sistema ortonormal*. Los vectores de un sistema ortonormal son linealmente independientes.

Si $\{\mathbf{q}_i\}$ es una base ortonormal, las coordendas de un vector $\mathbf{v}$ respecto de ella se pueden obtener como

$$v_i = \langle \mathbf{v}, \mathbf{q}_i \rangle$$

Para comprobarlo basta escribir $\mathbf{v}$ en términos de la base y multiplicar escalarmente ambos miembros de la ecuación por $\mathbf{q}_i$.

Las bases ortonormales facilitan el cálculo de productos escalares. Si (u_i) y (v_i) son, respectivamente, las coordenadas de los vectores $\mathbf{u}$ y $\mathbf{v}$ respecto de una base ortonormal,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} u_i \bar{v}_j,$$

de donde

$$\|\mathbf{u}\|^2 = \sum_{i=1}^{n} |u_i|^2.$$

Una matriz cuadrada es *unitaria* si sus columnas forman un sistema ortonormal. En el caso real se denominan *ortogonales*. Se demuestra fácilmente que las matrices unitarias tienen las siguientes propiedades:

1. Una matriz es unitaria si y sólo si su transpuesta conjugada coincide con su inversa.
2. Si Q es unitaria, Q^* y $Q^\top$ también lo son. Por tanto, sus filas conjugadas (y sin conjugar) forman un sistema ortonormal.
3. Una matriz es unitaria si y sólo si preserva el producto escalar.
4. El producto de dos matrices unitarias es también una matriz unitaria.

Ejercicios

1.10. Demostrar las propiedades enunciadas de las matrices unitarias.

1.11. (a) Demostrar que los vectores de un sistema de vectores ortogonales son linealmente independientes.
(b) Se considera una base ortonormal (es decir, formada por vectores ortogonales dos a dos y de norma unidad) $B = \{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$. Hallar la matriz R tal que $R\mathbf{x}$ son las coordenadas de $\mathbf{x}$ en la base B. (Indicación: desarrollar $\mathbf{x}$ en términos de los $\mathbf{q}_i$ y premultiplicar por $\mathbf{q}_k^*$).

1.7.3. Proyecciones ortogonales

Dado un subespacio E de un espacio vectorial de dimensión finita V el conjunto de vectores ortogonales a todos los de E es otro subespacio vectorial que se denomina *complemento ortogonal* de E, y se representa por $E^\perp$.

Consideramos el subespacio S generado por el sistema ortonormal $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$. Se define la *proyección ortogonal* de $\mathbf{v}$ sobre S como el vector $P_S(\mathbf{v}) \in S$ tal que la diferencia *(error de proyección)* $\mathbf{v} - P_S(\mathbf{v})$ es ortogonal a S. Veamos que $P_S(\mathbf{v})$ existe y es único. Si existe, tendrá una expresión de la forma

$$P_S(\mathbf{v}) = \alpha_1\mathbf{u}_1 + \ldots + \alpha_m\mathbf{u}_m.$$

Escribiendo

$$\mathbf{v} = P_S(\mathbf{v}) + (\mathbf{v} - P_S(\mathbf{v})) = \alpha_1\mathbf{u}_1 + \ldots + \alpha_m\mathbf{u}_m + (\mathbf{v} - P_S(\mathbf{v}))\,,$$

tenemos

$$\langle \mathbf{v}, \mathbf{u}_i \rangle = \alpha_i.$$

Por tanto

$$P_S(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \ldots + \langle \mathbf{v}, \mathbf{u}_m \rangle \mathbf{u}_m.$$

Por tanto la proyección ortogonal, si existe, es única. Pero falta comprobar que el vector así definido tiene la propiedad de ortogonalidad deseada. Para ello basta comprobar que el error de proyección es ortogonal a todos los vectores de la base:

$$\langle \mathbf{v} - P_S(\mathbf{v}), \mathbf{u}_i \rangle = \langle \mathbf{v}, \mathbf{u}_i \rangle - \langle P_S(\mathbf{v}), \mathbf{u}_i \rangle = \langle \mathbf{v}, \mathbf{u}_i \rangle - \langle \mathbf{v}, \mathbf{u}_i \rangle = 0.$$

Esta proyección tiene también la propiedad de asignar a cada vector $\mathbf{x}$ el vector de S más cercano. En efecto, como $\mathbf{x} - P_S\mathbf{x}$ es ortogonal a S, dado cualquier otro $\mathbf{y} \in S$,

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x} - P_S\mathbf{x} + P_S\mathbf{x} - y\|^2 = \|\mathbf{x} - P_S\mathbf{x}\|^2 + \|P_S\mathbf{x} - \mathbf{y}\|^2 \geq \|\mathbf{x} - P_S\mathbf{x}\|^2$$

donde se ha utilizado el teorema de Pitágoras.

En términos matriciales, como $\langle \mathbf{x}, \mathbf{u}_i \rangle = \mathbf{u}_i^* \mathbf{x}$, tenemos

$$P_S \mathbf{x} = \sum_{i=1}^{M} \mathbf{u}_i (\mathbf{u}_i^* \mathbf{x}) = \sum_{i=1}^{M} (\mathbf{u}_i \mathbf{u}_i^*) \mathbf{x} = \left(\sum_{i=1}^{M} \mathbf{u}_i \mathbf{u}_i^* \right) \mathbf{x} = \begin{pmatrix} \mathbf{u}_1 & \dots & \mathbf{v}_M \end{pmatrix} \begin{pmatrix} \mathbf{u}_1^* \\ \vdots \\ \mathbf{v}_M^* \end{pmatrix} \mathbf{x}.$$

De acuerdo con esta fórmula, las matrices de las proyecciones ortogonales son hermíticas. Recíprocamente, cualquier *matriz de proyección* (es decir, tal que $P^2 = P$) hermítica P es una matriz de proyección ortogonal, pues si $\mathbf{x} \in \ker P$, $\mathbf{y} = P\mathbf{z} \in \operatorname{Im} P$,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x} = \mathbf{z}^* P^* \mathbf{x} = \mathbf{z}^* P \mathbf{x} = \mathbf{z}^* \mathbf{0} = 0.$$

1.7.4. Ortogonalización de Gram-Schmidt

A partir de cualquier conjunto finito de vectores podemos obtener mediante el *algoritmo de Gram-Schmidt* un sistema ortonormal que genera el mismo subespacio que el conjunto inicial.

El algoritmo consiste en ir añadiendo vectores al sistema ortonormal tomando vectores sucesivos del conjunto inicial, restándoles su proyección ortogonal sobre el subespacio generado por los vectores ya almacenados en el sistema ortonormal y normalizando el resultado, cuando éste es no es nulo.

Utilizando ortogonalización de Gram-Schmidt se puede demostrar que cualquier sistema ortonormal puede completarse hasta formar un sistema ortonormal más grande que sea base de todo el espacio $\mathbf{C}^n$. De esta forma obtenemos una base adaptada al subespacio S definido por el sistema ortonormal inicial en el sentido de que sus primeros vectores son base ortonormal de S y los restantes base ortonormal de $S^\perp$. Utilizando una base de esta forma se comprueba que en un espacio de dimensión finita $(S^\perp)^\perp = S$ y que dimensión (S) + dimensión $(S^\perp) = n$.

1.8. Proyecciones en general

El concepto general de proyección no requiere de un producto escalar. Lo incluimos aquí para que se sepa por qué las proyecciones ortogonales tienen necesidad de apellido.

Dos subespacios E y F de V son *complementarios* si verifican
(a) $\operatorname{span}\{E \cup F\} = V$.
(b) $E \cap F = \{0\}$.

La propiedad (a) significa que cualquier vector de V se puede escribir como

$$\mathbf{x} = \mathbf{x}_E + \mathbf{x}_F,\ \mathbf{x}_E \in E,\ \mathbf{x}_F \in F.$$

Además, como consecuencia de la propiedad (b), esta descomposición es única (ejercicio 1.12).

Si tomamos una base $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ en E y una base $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ en F, un vector $\mathbf{x}$ de V se podrá escribir como

$$\mathbf{x} = \underbrace{x_1\mathbf{u}_1 + \ldots + x_m\mathbf{v}_m}_{\mathbf{x}_E} + \underbrace{x'_1\mathbf{v}_1 + \ldots + x'_n\mathbf{v}_n}_{\mathbf{x}_F}.$$

Luego la unión de las bases es una base de V y la aplicaciones que asignan a cada $\mathbf{x}$ su vectores $\mathbf{x}_E$ y $\mathbf{x}_F$ son aplicaciones lineales, con matrices muy sencillas y muy fáciles de calcular a partir de la fórmula anterior). La aplicación lineal P_E definida por $P_E\mathbf{x} = \mathbf{x}_E$ se denomina *proyección sobre E paralela a F*, y la definida por $P_F\mathbf{x} = \mathbf{x}_F$ se denomina *proyección sobre F paralela a E*. Obviamente tenemos la propiedad

$$P_E + P_F = I. \tag{1.5}$$

Ejercicios

1.12. Demostrar que E y F son subespacios complementarios del espacio V, todo vector de V se puede escribir de forma única como suma de un vector de E más otro de F. Indicación: Utilizar la propiedad (b) de la definición.

1.9. Normas

Una *norma* en un espacio vectorial V es una aplicación de V en $\mathbf{R}^+$ y se verifica
1) $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
2) $\|\alpha\mathbf{u}\| = |\alpha|\|\mathbf{u}\|$
3) $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = 0$
Definiendo $d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|$ tenemos un espacio métrico.

Se verifica

$$\mathbf{x}_n \to \mathbf{x},\ \mathbf{y}_n \to \mathbf{y} \Rightarrow \alpha\mathbf{x}_n + \beta\mathbf{y}_n \to \alpha\mathbf{x} + \beta\mathbf{y}.$$

Dos normas $\|\cdot\|_1$, $\|\cdot\|_2$ son *equivalentes* si existen constantes c_1, c_2 tales que para todo vector x

$$c_1\|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \leq c_2\|\mathbf{x}\|_1.$$

Si una secuencia converge a un límite en una norma, converge al mismo límite también en las normas equivalentes a ella.

En $\mathbf{C}^n$ se definen las normas p, $p \geq 1$, y la norma ∞ como

$$\|\mathbf{x}\|_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}$$
$$\|\mathbf{x}\|_\infty = \max_{k=1,\ldots,n} |x_k|.$$

Una norma p y la norma ∞ son equivalentes en $\mathbf{C}^n$:

$$\|\mathbf{x}\|_p^p = \sum_{k=1}^{n} |x_k|^p \leq n\,\|x\|_\infty^p\,,$$
$$\|\mathbf{x}\|_\infty^p \leq \sum_{k=1}^{n} |x_k|^p = \|x\|_p\,,$$

luego

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_p \leq n^{1/p}\,\|\mathbf{x}\|_\infty\,.$$

De hecho se puede demostrar que en un espacio de dimensión finita dos normas cualesquiera son equivalentes.

Una norma deriva de un producto escalar si y sólo si verifica la *igualdad del paralelogramo:*

$$2\left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2\right) = \|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2\,.$$

Entonces el producto escalar se puede recuperar a partir de la norma mediante la fórmula

$$\langle\mathbf{x},\mathbf{y}\rangle = \frac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 + i\,\|\mathbf{x}+i\mathbf{y}\|^2 - i\,\|\mathbf{x}-i\mathbf{y}\|^2\right),$$

que se denomina *identidad de polarización.*

Capítulo 2

Matrices hermíticas

2.1. Definición y propiedades básicas

Como hemos indicado anteriormente, las matrices *hermíticas* son las matrices que coinciden con su traspuesta conjugada. Si son reales, equivalen a las simétricas. Las siguientes propiedades son fáciles de demostrar:
(1) Tienen autovalores reales.
(2) Autovectores con distintos autovalores son ortogonales.

2.2. Diagonalización de matrices hermíticas

Las matrices hermíticas tienen bases ortonormales de autovectores y, por tanto, son unitariamente diagonalizables.

Vamos a demostrar esto último por inducción. Para matrices 1×1 es trivialmente cierto, porque todas las matrices de este tamaño son diagonales. Suponemos que toda matriz hermítica $n \times n$ se puede factorizar como

$$A = QDQ^* \tag{2.1}$$

con Q unitaria y D diagonal real. Veamos que si A es hermítica $(n+1)\times(n+1)$ también existe una factorización análoga. Para ello construimos una matriz unitaria $Q_1 = (\mathbf{q}_1, \ldots, \mathbf{q}_{n+1})$ de forma que su primera columna, $\mathbf{q}_1$, sea un autovector de A con autovector λ. Entonces

$$Q_1^* A Q_1 = Q_1^* \begin{pmatrix} A\mathbf{q}_1 & \ldots & A\mathbf{q}_{n+1} \end{pmatrix} = Q_1^* \begin{pmatrix} \lambda\mathbf{q}_1 & \ldots & A\mathbf{q}_{n+1} \end{pmatrix} = \begin{pmatrix} \lambda & \mathbf{v}^* \\ & B \end{pmatrix}.$$

Pero $\mathbf{v} = \mathbf{0}$ por ser la matriz simétrica. Como B es $n \times n$, $B = Q_2 D Q_2^*$ para cierta D diagonal real y Q_2 unitaria, luego

$$Q_1^* A Q_1 = \begin{pmatrix} \lambda & \\ & B \end{pmatrix} = \begin{pmatrix} \lambda & \\ & Q_2 D Q_2^* \end{pmatrix} = \begin{pmatrix} 1 & \\ & Q_2 \end{pmatrix} \begin{pmatrix} \lambda & \\ & D \end{pmatrix} \begin{pmatrix} 1 & \\ & Q_2 * \end{pmatrix}$$

de donde despejando A obtenemos la factorización buscada (2.1), que garantiza, de acuerdo con 1.6.2, que existe una base ortonormal formada por autovectores de A, concretamente por las columnas de Q.

Veamos ahora la unicidad de la factorización. Supongamos que tenemos dos factorizaciones

$$A = Q_1 D_1 Q_1^* = Q_2 D_2 Q_2^*$$

con $D_i = \operatorname{diag}(\lambda_1^{(i)}, \ldots, \lambda_n^{(i)})$, $\lambda_1^{(i)} \geq \ldots \geq \lambda_n^{(i)}$. Entonces, como los elementos de la matriz diagonal son los autovalores de A repetidos según su multiplicidad, ambas matrices D_i deben ser iguales, y $\lambda_k^{(1)} = \lambda_k^{(2)} = \lambda_k$.

En cuanto a las matrices $Q_i = (\mathbf{q}_1^{(i)}, \ldots, \mathbf{q}_n^{(i)})$, los $\mathbf{q}_k^{(i)}$ deben ser autovectores unitarios con autovalor λ_k. La relación más general es que los conjuntos de columnas de cada Q_i asociados al mismo autovalor son bases ortonormales del mismo subespacio (el subespacio asociado al autovalor). En particular, las columnas asociadas a un autovalor simple λ_k verificarán $\mathbf{q}_k^{(1)} = c_k \mathbf{q}_k^{(2)}$, con $|c_k| = 1$.

2.3. Clasificación de matrices hermíticas

Las matrices hermíticas no nulas se clasifican en los tipos disjuntos siguientes:
(1) Definidas positivas: Si para todo $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^* A \mathbf{x} > 0$.
(2) Semidefinidas positivas: Si para algunos $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^* A \mathbf{x} > 0$ y para otros $\mathbf{x}^* A \mathbf{x} = 0$.
(3) Definidas negativas: Si para todo $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^* A \mathbf{x} < 0$.
(4) Semidefinidas negativas: Si para algunos $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^* A \mathbf{x} < 0$ y para otros $\mathbf{x}^* A \mathbf{x} = 0$.
(5) Indefinidas: Si para algunos $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^* A \mathbf{x} < 0$, y para otros $\mathbf{x}^* A \mathbf{x} > 0$.

Se demuestra que estos tipos se corresponden respectivamente con los casos
(1) Todos los autovalores positivos.
(2) Algunos autovalores positivos y otros nulos.
(3) Todos los autovalores negativos.
(4) Algunos autovalores negativos y otros nulos.
(5) Algunos autovalores positivos y otros negativos.

Ejercicios

2.1. Demostrar las siguientes propiedades de las matrices hermíticas:

(a) Tienen autovalores reales.
(b) Autovectores con distintos autovalores son ortogonales.

2.2. Demostrar las siguientes propiedades de las matrices hermíticas:
(a) Si A es hermítica, VAV^* también lo es.
(b) Si A es hermítica y triangular, A es diagonal.
(c) Demostrar que toda matriz hermítica es unitariamente diagonalizable (indicación: utilizar la factorización de Schur (ejercicio 2.6) (no utilizar el ejercicio 2.7)).

2.3. Demostrar la correspondencia indicada en el texto entre los tipos de matrices hermíticas y los signos de sus autovalores.

2.4. Demostrar que las matrices de la forma $A = B^*B$ son definidas positivas o semidefinidas positivas.

2.5. (a) ¿Cuáles son los autovalores de una matriz triangular?
(b) ¿Qué relación existe entre los autovectores y los autovalores de A y los de A^{-1}? ¿Y entre los de A y los de $A + \mu I$?

Capítulo 3

Análisis de componentes principales

3.1. Resultado fundamental

Si $\mathbf{X}$ es una variable aleatoria real de dimensión n de media nula con matriz de covarianzas $\Sigma_{\mathbf{X}} = E[\mathbf{X}\mathbf{X}^\top]$, definimos su *varianza total* como

$$V_T[\mathbf{X}] = E[\|\mathbf{X}\|^2] = \mathrm{tr}(\Sigma_{\mathbf{X}}).$$

Queremos encontrar un subespacio V de dimensión $m < n$ tal que la varianza total de la variable aleatoria $\mathbf{Y} = P_V\mathbf{X}$ sea máxima.

Veamos que este subespacio es el generado por los m autovectores de $\Sigma_{\mathbf{X}}$ de mayores autovalores.

Para demostrar esta propiedad obtenemos en primer lugar una expresión conveniente de la cantidad a maximizar. Si los vectores $\{\mathbf{v}_1, ..., \mathbf{v}_m\}$ son una base ortonormal del subespacio buscado V, la proyección ortogonal sobre este subespacio vendrá dada por

$$\mathbf{Y} = P_V\mathbf{X},\ P_V = \underbrace{\begin{pmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_m\end{pmatrix}}_{\equiv B} \underbrace{\begin{pmatrix}\mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_m^T\end{pmatrix}}_{\equiv B^\top}$$

Tenemos maximizar

$$E[\|\mathbf{Y}\|^2] = E[\mathbf{Y}^\top\mathbf{Y}] = E[\mathrm{tr}(\mathbf{Y}^\top\mathbf{Y})] = E[\mathrm{tr}(\mathbf{Y}\mathbf{Y}^\top)] = \mathrm{tr}(E[\mathbf{Y}\mathbf{Y}^\top]) = \mathrm{tr}\,\Sigma_Y,$$
$$\text{donde } \Sigma_Y \equiv E[\mathbf{Y}\mathbf{Y}^\top] = E[P_V\mathbf{X}\mathbf{X}^\top P_V^\top] = P_V E[\mathbf{X}\mathbf{X}^\top]P_V^\top = P_V\Sigma P_V^\top.$$

Como Σ es una matriz simétrica definida positiva, existe una base ortonormal formada por autovectores suyos. Sean $(\mathbf{u}_i)_{i=1,\dots,n}$ estos autovectores, con autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$. Vamos a escribir los vectores que buscamos en términos de esta base:

$$\mathbf{v}_i = \sum_{j=1}^{n} v_{ij}\mathbf{u}_j.$$

Matricialmente

$$B = \begin{pmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_m\end{pmatrix} = \underbrace{\begin{pmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_n\end{pmatrix}}_{U} \underbrace{\begin{pmatrix} v_{11} & \dots & v_{m1} \\ & \dots & \\ v_{1n} & \dots & v_{mn}\end{pmatrix}}_{A}$$

Como los $\mathbf{u}_i$ son un sistema ortonormal real, U es ortogonal. Como hemos tomado la base $\mathbf{v}_i$ ortonormal, las columnas de la matriz A también son un sistema ortonormal o, lo que es lo mismo, A está formada por m columnas de una matriz ortogonal:

$$A^\top A = B^\top U U^\top B = B^\top B = I_m$$

Vamos a expresar la cantidad a maximizar en función de estas matrices:

$$E[\|\mathbf{Y}\|^2] = \operatorname{tr}(\Sigma_Y) = \operatorname{tr}(P_V \Sigma P_V^\top) = \operatorname{tr}(UAA^\top \underbrace{U^\top \Sigma U}_{\Lambda} AA^\top U^\top) = \operatorname{tr}(A^\top \Lambda A)$$

$$= \operatorname{tr}(\Lambda AA^\top) = \sum_{j=1}^{n} \lambda_j \mathbf{a}_j^\top \mathbf{a}_j = \sum_{j=1}^{n} \lambda_j \|\mathbf{a}_j\|^2,$$

donde los $\mathbf{a}_j^\top$ son las filas de A.

Podemos escribir esta cantidad como

$$\sum_{j=1}^{n} \lambda_j c_j, \tag{3.1}$$

$$\text{donde } \; c_j = \|\mathbf{a}_j\|^2 = \sum_{i=1}^{m} v_{ij}^2. \tag{3.2}$$

Cada c_j son suma de los m primeros coeficientes al cuadrado de la fila j de una matriz ortogonal, luego

$$0 \leq c_j \leq 1, \tag{3.3}$$

y la suma de los c_j es la suma de los cuadrados de los coeficientes de las m primeras filas de la matriz, y por tanto,

$$\sum_{j=1}^{n} c_j = m. \tag{3.4}$$

Por tanto podemos plantearnos maximizar (3.1) con las restricciones (3.3) y (3.4) y ver si esta solución se corresponde con alguna elección de coeficientes de vectores ortonormales v_{ij}.

Con $v_{ij} = \delta_{ij}$, lo que corresponde a tomar los $\mathbf{v}_i = \mathbf{u}_i$, $i = 1, \ldots, m$, tenemos $c_1 = \ldots = c_m = 1$, $c_{m+1} = \ldots = c_n = 0$, luego

$$\sum_{j=1}^{n} \lambda_j c_j = \sum_{j=1}^{m} \lambda_j$$

La función que queremos maximizar no puede tomar valores mayores que estos, pues para una elección arbitraria de coeficientes c_i que verifiquen las restricciones (3.3) y (3.4),

$$\begin{aligned}
\sum_{j=1}^{n} \lambda_j c_j - \sum_{j=1}^{m} \lambda_j &= \sum_{j=1}^{m} \lambda_j (c_j - 1) + \sum_{j=m+1}^{n} \lambda_j c_j \\
\leq \sum_{j=1}^{m} \lambda_j (c_j - 1) + \lambda_{m+1} & \underbrace{\sum_{j=m+1}^{n} c_j}_{} \\
&= m - \sum_{j=1}^{m} c_j = \sum_{j=1}^{m} (1 - c_j) \\
= \sum_{j=1}^{m} \underbrace{(c_1 - 1)}_{\leq 0} \underbrace{(\lambda_j - \lambda_{m+1})}_{\geq 0} &\leq 0.
\end{aligned}$$

3.2. Variaciones sobre el tema

3.2.1. Minimización de la varianza total

De forma análoga podríamos comprobar que el subespacio U de dimensión r que minimiza $E[\|P_U \mathbf{X}\|^2]$ es el generado por los autovectores de Σ con menores autovalores, y entonces $E[\|P_U \mathbf{X}\|^2] = \sum_{i=n-r+1}^{n} \lambda_i$.

Visto de otra forma, como $P_{W^\perp} = I - P_W$,

$$\begin{aligned}
\mathbf{X} = P_W \mathbf{X} + P_{W^\perp} \mathbf{X} \Rightarrow \|\mathbf{X}\|^2 = \|P_W \mathbf{X}\|^2 + \|P_{W^\perp} \mathbf{X}\|^2 \\
\Rightarrow E[\|\mathbf{X}\|^2] = E[\|P_W \mathbf{X}\|^2] + E[\|P_{W^\perp} \mathbf{X}\|^2],
\end{aligned}$$

luego si W maximiza $E[\|P_W\mathbf{X}\|^2]$ entre los subespacios de dimensión m, $U = W^\perp$ minimiza $E[\|P_U\mathbf{X}\|^2]$ entre los subespacios de dimensión $n-m$. Y como W es el subespacio generado por los autovectores de mayores autovalores, $W^\perp$ es el subespacio generado por los autovectores restantes, que son los de menores autovalores.

3.2.2. Obtención de ecuaciones lineales

También podemos estar interesados en hallar las m ecuaciones lineales que satisfacen un cierto conjunto de datos, de los que tenemos observaciones ruidosas que modelamos mediante la variable $\mathbf{X} \in \mathbf{R}^n$, de matriz de covarianzas Σ.

Los coeficientes de estas ecuaciones son un subespacio vectorial W de $\mathbf{R}^n$. Vamos a buscar una base ortonormal $\mathbf{u}_i$ de este subespacio. Si escribimos estas ecuaciones como

$$\mathbf{u}_i^\top \mathbf{X} = 0,\ i = 1, \ldots, m,$$

parece razonable intentar minimizar

$$E\left[\sum_{i=1}^{m} (\mathbf{u}_i^\top \mathbf{X})^2\right] = E[\|P_W\mathbf{X}\|^2].$$

Así que este es el problema de análisis de componentes principales cambiando la maximización por minimización de la función. La solución corresponderá por tanto a los m autovectores de menores autovalores.

3.2.3. Análisis de componentes principales y estimación de máxima verosimilitud

Observemos en primer lugar que $\|P_{W^\perp}\mathbf{X}\|$ es la distancia $d(\mathbf{X}, W)$ del vector X al subespacio W (porque es la norma del error de proyección de $\mathbf{X}$ sobre W, que es la mínima distancia entre $\mathbf{X}$ y un punto de W).

Por tanto el subespacio W generado por los autovalores de mayores autovectores minimiza $E[d^2(\mathbf{X}, W)]$. Esto permite dar al análisis de componentes principales una interpretación en términos de estimación de máxima verosimilitud.

Supongamos un conjunto de puntos $\hat{\mathbf{X}}_i \in \mathbf{R}^n, i = 1, \ldots, r$ que sabemos que están en un subespacio vectorial W, desconocido, de dimensión conocida m, y de los que tenemos observaciones ruidosas $\mathbf{X}_i = \hat{\mathbf{X}}_i + \mathbf{Z}_i$, donde los $\mathbf{Z}_i$ son realizaciones de variables aleatorias independientes gaussianas de media nula y componentes independientes de varianza σ^2.

Entonces la densidad de probabilidad de la observación condicionada a los valores de los parámetros desconocidos es

$$f(\mathbf{X}_1,\ldots,\mathbf{X}_r|\hat{\mathbf{X}}_1,\ldots,\hat{\mathbf{X}}_r) = k\exp -\frac{\sum_{i=1}^r \|\mathbf{X}_i-\hat{\mathbf{X}}_i\|^2}{2\sigma^2},$$

luego la estimación de máxima verosimilitud consistirá en encontrar unos $\hat{\mathbf{X}}_i$ que verifiquen la restricción, es decir, que estén en un subespacio de dimensión m, y que minimicen $\sum_{i=1}^r \|\mathbf{X}_i-\hat{\mathbf{X}}_i\|^2$.

Si suponemos que el subespacio es W, la solución sería $\tilde{\mathbf{X}}_i = P_W\mathbf{X}_i$, y el valor de la función de coste sería

$$k\exp -\frac{\sum_{i=1}^r d^2(\mathbf{X}_i,W)}{2\sigma^2}.$$

En consecuencia, el W óptimo será el proporcionado por el análisis de componentes principales.

3.3. Aplicación a conjuntos discretos de datos

Si disponemos de un conjunto discreto de M datos $\mathbf{X}_i$ de dimensión n y queremos encontrar un subespacio de dimensión inferior sobre el que proyectarlos conservando la mayor parte de la energía podemos aplicar la teoría que acabamos de desarrollar sin más que identificar este conjunto de datos con observaciones de una variable aleatoria cuya matriz de covarianzas podemos estimar por

$$\Sigma = \frac{1}{M}\sum_{i=1}^M \mathbf{X}_i\mathbf{X}_i^\top = \frac{1}{M}AA^*.$$

donde

$$A = \begin{pmatrix}\mathbf{X}_1 & \ldots & \mathbf{X}_M\end{pmatrix}.$$

Como AA^* es una matriz $n\times n$, si la dimensión n del conjunto de datos es muy grande el cálculo de autovectores puede ser muy costoso. Sin embargo, si $M<n$ este cálculo se puede simplificar basándose en el cálculo de autovalores y autovectores de A^*A, que es $M\times M$, puesto que, como es fácil comprobar, si $\mathbf{u}$ es autovector de A^*A, $A\mathbf{u}$ es autovector de AA^* con el mismo autovalor. Esta correspondencia entre autovectores preserva la ortogonalidad y proporciona todos los autovectores de AA^* con autovalor no nulo.

Si $U=(\mathbf{u}_1,\ldots,\mathbf{u}_m)$ y $\Lambda=\operatorname{diag}(\lambda_1,\ldots,\lambda_m)$ son las matrices de autovectores unitarios y autovalores de A^*A, la matriz de $V=(\mathbf{v}_1,\ldots,\mathbf{v}_m)$ de autovectores unitarios de AA^* es, utilizando que $\mathbf{v}_i = A\mathbf{u}_i/\sqrt{\lambda_i}$,

$$V = AU\Lambda^{-1/2},$$

con lo que las coordenadas de los datos en la base de los $\mathbf{v}_i$ son

$$V^*A = \Lambda^{-1/2}U^*A^*A = \Lambda^{-1/2}U^*U\Lambda U^* = \Lambda^{1/2}U^*.$$

Por tanto los autovectores de A^*A nos dan directamente las coordenadas de los datos en la base de componentes principales.

Ejercicios

3.1. Utilizando el programa para calcular el ACP de bloques de una imagen, trabajar con una imagen de 640×480 y con bloques de 32×32 y estudiar cómo varía la energía contenida en n coeficientes en función del número de componentes con que nos quedamos. Explicar por qué la gráfica termina siendo constante.

3.4. Análisis discriminante lineal

Ahora consideramos datos de distintas clases con distintas distribuciones. Concretamente, tenemos C clases, la probabilidad de que un dato pertenezca a la clase c es P_c y los datos de esta clase presentan una media μ_c y una matriz de covarianzas Σ_c. Notamos como μ la media global de los datos.

Para un conjunto de datos como este se definen las siguientes *matrices de dispersión:*

- *Matriz de dispersión total:*

$$S_T = E\left[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^T\right].$$

- *Matriz de dispersión intra-clases*

$$S_W = \sum_{c=1}^{C} P_c E\left[(\mathbf{X}-\mu_c)(\mathbf{X}-\mu_c)^T|c\right] = \sum_{c=1}^{C} P_c \Sigma_c.$$

- *Matriz de dispersión entre clases*

$$S_B = \sum_{c=1}^{C} P_c(\mu_c-\mu)(\mu_c-\mu)^T.$$

Se comprueba (ejercicio) que

$$S_T = S_W + S_B.$$

Queremos encontrar una dirección sobre la que proyectar los datos tal que maximice la varianza entre clases de los datos proyectados, con la restricción de que la varianza intra-clases de los datos proyectados sea un valor constante.

Si $\mathbf{v}$ es un vector unitario que define la dirección que buscamos, los datos proyectados corresponderán con la variable aleatoria escalar $\mathbf{v}^T\mathbf{X}$. Los datos de cada clase se proyectarán sobre datos con media $\mathbf{v}^T\mu_c$ y varianza $\mathbf{v}^T\Sigma_c\mathbf{v}$, y la media global será $\mathbf{v}^T\mu$.

La varianza entre clases será por tanto

$$\sigma_B^2 = \sum_c P_c(\mathbf{v}^T\mu_c)(\mathbf{v}^T\mu_c)^T = \mathbf{v}^T\left(\sum_c P_c\mu_c\mu_c^T\right)\mathbf{v} = \mathbf{v}^T S_B\mathbf{v},$$

y análogamente la varianza intraclases será **(ejercicio)**

$$\sigma_W^2 = \mathbf{v}^T S_W\mathbf{v}.$$

El problema es por tanto

$$\text{maximizar}_{\mathbf{v}}\ \mathbf{v}^t S_B \mathbf{v} \text{ con la restricción } \mathbf{v}^T S_W \mathbf{v} = \text{cte},$$

cuya lagrangiana es

$$\mathbf{v}^t S_B \mathbf{v} + \lambda \mathbf{v}^T S_W \mathbf{v}.$$

Igualando el gradiente a cero queda

$$S_B \mathbf{v} = \lambda S_W \mathbf{v} \Leftrightarrow S_W^{-1} S_B \mathbf{v} = \lambda \mathbf{v},$$

luego $\mathbf{v}$ es un autovector de $S_W^{-1} S_B$.

Este problema de autovalores se puede transformar en un problema de autovalores de matriz simétrica mediante la transformación $\mathbf{v}' = S_B^{1/2} \mathbf{v}$ (ejercicio).

3.5. Escalado multidimensional

Partiendo de las distancias d_{ij} entre cada par de un conjunto de datos queremos asignar a cada dato un vector de coordenadas en $\mathbf{R}^n$, con n lo más pequeño posible, de forma que se preserven las distancias entre los datos.

Para ello, partimos de la matriz

$$A = (a_{ij}),\ a_{ij} = -\frac{1}{2} d_{ij}^2$$

y, definiendo la matriz

$$H = I - \frac{1}{n} \mathbf{1}\mathbf{1}^T$$

calculamos la matriz

$$B = HAH.$$

El resultado que nos resuelve el problema es el siguiente [5].

- El problema tiene solución si y sólo si B es (semi)definida positiva.
- En ese caso la solución del problema viene dada por la factorización

  $$B = UDU^T = RR^T,\ R = UD^{1/2}.$$

  Las coordenadas que buscamos son las filas de R y corresponden a un conjunto de datos de media nula.

Capítulo 4

Descomposición en valores singulares

4.1. Teorema de descomposición en valores singulares

El *teorema de descomposición en valores singulares (singular value decomposition (SVD))* asegura que toda matriz admite una descomposición de la forma

$$A = USV^*$$

donde U y V son matrices ortogonales y S es diagonal de las mismas dimensiones que A y con elementos reales no negativos que se denominan *valores singulares* de A y son únicos. Además, si un valor singular σ_i no se repite, los vectores correspondientes de las columnas de U y V son únicos salvo por el signo. Las columnas de U y V se denominan respectivamente *vectores principales por la izquierda y por la derecha* de A.

Consideramos la aplicación f de $\mathbf{C}^n$ en $\mathbf{C}^m$ de matriz A $(m \times n)$. Tomamos una una base ortonormal $\{\mathbf{v}_i\}_{i=1,\dots,n}$ del espacio de partida formada por autovectores de la matriz hermítica A^*A y denominamos σ_i^2 a sus autovalores asociados (que sabemos que son reales no negativos. Consideramos los $\mathbf{v}_i$ ordenados de forma que los σ_i son decrecientes.

Para cada $\sigma_i \neq 0$ definimos $\mathbf{u}_i = A\mathbf{v}_i/\sigma_i$. Estos vectores son también un conjunto ortonormal:

$$\mathbf{u}_i^*\mathbf{u}_j = \frac{1}{\sigma_i\sigma_j}\mathbf{v}_i^*A^*A\mathbf{v}_j = \frac{1}{\sigma_i\sigma_j}\mathbf{v}_i^*(A^*A\mathbf{v}_j) = \frac{1}{\sigma_i\sigma_j}\mathbf{v}_i^*(\sigma_j^2\mathbf{v}_j) = \delta_{ij}.$$

Añadiendo vectores ortonormales a este conjunto hasta completar una base ortonormal de $\mathbf{C}^m$ obtenemos los vectores $\{\mathbf{u}_i\}_{i=1\dots,m}$. La aplicación en estas bases es diagonal, pues

$$f(\mathbf{v}_i) = \sigma_i\mathbf{u}_i.$$

Por tanto la matriz A se puede escribir como el producto de las matrices de cambio de base por una matriz diagonal:

$$A = USV^*.$$

Para comprobar la unicidad de los elementos de la SVD, observamos que si $A = USV^*$, $A^*A = VS^\top SV$ es una diagonalización de A^*A. De ahí se desprende la unicidad de los σ_i y que si un σ_i no se repite, su $\mathbf{v}_i$, por ser el autovector correspondiente, es único salvo por un factor de escala c_i de módulo unidad. Además $\mathbf{u}_i = A\mathbf{v}_i/\sigma_i$, luego $\mathbf{u}_i$ también está determinado unívocamente salvo por la misma constante c_i.

Ejercicios

4.1. Calcular la SVD de las matrices $A = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ y representar gráficamente los vectores obtenidos sobre la circunferencia unidad y su imagen.

4.2. ¿Qué relación existe entre la SVD de una matriz hermítica y su diagonalización?

4.2. Norma de una matriz

Entre las posibles normas que se pueden tomar en el conjunto de las matrices $m \times n$ consideraremos la *norma inducida* por las normas euclídeas en los espacios inicial y final, es decir, definimos

$$\|A\| = \sup_{\mathbf{x}\in\mathbf{C}^n} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|.$$

Por tanto $\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$. Es fácil comprobar que

$$\|A\| = \sigma_1,$$

donde σ_1 es el mayor valor singular de A. Si A es regular, $\|A^{-1}\| = 1/\sigma_n$, donde σ_n es el menor valor singular de A. Otra propiedad inmediata es que $\|AB\| \leq \|A\|\|B\|$.

En ocasiones se utilizan otras normas de matrices, como la *norma de Frobenius*, definida por

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2} = \sqrt{\mathrm{tr}(AA^*)}.$$

Cuando hagamos referencia a una norma sin indicar a cuál nos referiremos a la norma inducida por la norma euclídea.

Es inmediato que estas normas de una matriz son invariantes por multiplicación de la matriz por cualquier lado por una matriz unitaria. Utilizando esta propiedad,

$$\|A\|_F = \sqrt{\sum_i^r \sigma_i^2}.$$

Como son dos normas sobre un espacio de dimensión finita, son equivalentes. De hecho, de la relación anterior,

$$\|A\| \leq \|A\|_F \leq \sqrt{r}\|A\|$$

La norma inducida no deriva de un producto escalar, pero la norma de Frobenius claramente corresponde al producto $\langle A, B\rangle = \mathrm{tr}(AB^*)$.

4.3. Aplicaciones y propiedades de la SVD

4.3.1. Núcleo y rango de una matriz

La SVD nos proporciona bases ortonormales del núcleo y la imagen de una matriz. Una base ortonormal del núcleo es la dada por los vectores singulares por la derecha que corresponden con valores singulares nulos. Una base ortonormal de la imagen es la dada por los vectores singulares por la izquierda que corresponden a valores singulares no nulos.

4.3.2. Aproximación de una matriz por otra de rango inferior

Dada la matriz A, $m \times n$, de rango $r \leq \mín\{m, n\}$, podemos escribirla como suma de matrices de rango uno en términos de su SVD:

$$A = \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^*, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

La mejor aproximación de rango $q < r$ de A en términos de la norma inducida es

$$A_q = \sum_{k=1}^{q} \sigma_k \mathbf{u}_k \mathbf{v}_k^*$$

y

$$\|A - A_q\| = \sigma_{q+1}.$$

Lo demostramos por reducción al absurdo. Supongamos que existiera una matriz B $m \times n$ de rango menor o igual que q tal que $\|A - B\| < \sigma_{q+1}$.

El núcleo de B es un subespacio W de dimensión $n - \text{rank}(B) \geq n - q$, y para $\mathbf{w} \in W$

$$\|A\mathbf{w}\| = \|(A - B)\mathbf{w}\| \leq \|A - B\|\|\mathbf{w}\| < \sigma_{q+1}\|\mathbf{w}\|.$$

Pero existe un subespacio V de dimensión $q+1$ (el generado por $\mathbf{v}_1, \ldots, \mathbf{v}_{q+1}$) en el que para todo vector $\mathbf{v} = \alpha_1\mathbf{v}_1 + \ldots + \alpha_{q+1}\mathbf{v}_{q+1} \in V$

$$\|A\mathbf{v}\|^2 = (\sigma_1\alpha_1)^2 + \ldots + (\sigma_{q+1}\alpha_{q+1})^2 \geq \sigma_{q+1}^2\|\mathbf{v}\|^2$$
$$\Rightarrow \|A\mathbf{v}\| \geq \sigma_{q+1}\|\mathbf{v}\|.$$

Pero como W tiene dimensión mayor o igual que $n - q$ y V tiene dimensión $q + 1$, W y V tienen que cortarse no trivialmente, con lo que existirá algún vector no nulo $\mathbf{x}$ para el que al mismo tiempo

$$\|A\mathbf{x}\| < \sigma_{q+1}\|\mathbf{x}\| \text{ y } \|A\mathbf{x}\| \geq \sigma_{q+1}\|\mathbf{x}\|,$$

que es la contradicción que buscamos.

El hecho de que $\|A - A_q\| = \sigma_{q+1}$ se comprueba directamente.

Ejercicios

4.3. Utilizando la SVD, demostrar
(a) El rango de una matriz es igual al de su traspuesta.
(b) La dimensión de la imagen más la dimensión del núcleo es igual al número de columnas (indicar qué vectores son base ortonormal de su núcleo y qué vectores lo son de su imagen).
(c) A^*A y AA^* tienen los mismos autovalores.
(d) Si una matriz cuadrada preserva la norma, es unitaria.

4.4. Se define la *norma* $\|A\|$ de una matriz A como el máximo valor de $\|A\mathbf{x}\|$ cuando $\|x\| = 1$. Demostrar que $\|A\| = \sigma_1$ (mayor valor singular de A).

4.5. Demostrar que si A es una matriz cuadrada regular $n \times n$, la matriz singular más cercana es la que se obtiene sustituyendo el menor valor singular σ_n por cero en la SVD de A. Indicación: La matriz propuesta está a distancia σ_n de A. Por la definición de norma, para demostrar que $\|A - B\| \geq \sigma_n$ basta encontrar un vector unitario v tal que $\|(A - B)v\|_2 \geq \sigma_n$. Tomar un vector del núcleo de B.

4.6. Indicar, utilizando la SVD de $A = USV^*$, cómo se puede obtener la diagonalización de AA^* si se conoce la diagonalización de A^*A. Indicación: Tener en cuenta la relación $AV = US$. Teniendo en cuenta la sección 3.3, explicar cómo se podría aplicar de una forma práctica ACP al caso en el que se dispone en un banco de datos de 1000 imágenes de caras de personas, cada una de 1000×1000 píxeles y se desea proyectar los datos de cada imagen sobre un espacio de dimensión 50 antes de aplicar un algoritmo de reconocimiento automático.

4.7. Obtener mediante SVD bases ortonormales del núcleo y la imagen de las matrices

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix},\ B = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix},\ C = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Capítulo 5

Problemas de mínimos cuadrados

5.1. El problema de mínimos cuadrados

Si A es una matriz rectangular $m \times n$, $m > n$, de rango máximo, para un $\mathbf{b} \in \mathbf{C}^n$ dado la ecuación

$$A\mathbf{x} = \mathbf{b} \tag{5.1}$$

no tendrá en general solución. Solo la tendrá cuando $\mathbf{b}$ sea combinación lineal de las columnas de A. Sin embargo siempre podemos buscar el vector $\mathbf{x}$ que más se acerque a satisfacer la ecuación, es decir, que minimice

$$\|A\mathbf{x} - \mathbf{b}\|^2.$$

Esta es la solución del problema (5.1) en el sentido de mínimos cuadrados.

Buscamos por tanto un $\mathbf{x}$ que dé lugar a una combinación lineal de las columnas de A, $\{\mathbf{a}_i\}$, que esté lo más cerca posible de $\mathbf{b}$. Esta combinación lineal no es otra cosa que la proyección ortogonal de $\mathbf{b}$ sobre la imagen de A. Por tanto el $\mathbf{x}$ buscado verifica $\mathbf{b} - A\mathbf{x} \perp \mathbf{a}_i$, y de aquí obtenemos $\mathbf{x}$

$$A^*(\mathbf{b} - A\mathbf{x}) = 0 \Rightarrow A^*A\mathbf{x} = A^*\mathbf{b}. \tag{5.2}$$

Las ecuaciones (5.2) se denominan *ecuaciones normales*. Como A es de rango máximo, A^*A es invertible (como se ve, p. ej., usando SVD), luego

$$\mathbf{x} = (A^*A)^{-1}A^*\mathbf{b}.$$

La matriz

$$A^+ = (A^*A)^{-1}A^* \tag{5.3}$$

se denomina *pseudoinversa (de Moore-Penrose)* de A.

La aproximación a $\mathbf{b}$ que obtenemos es

$$A\mathbf{x} = A(A^*A)^{-1}A^*\mathbf{b}.$$

Así que de paso hemos obtenido que $P = A(A^*A)^{-1}A^*$ es la matriz de la proyección ortogonal sobre la imagen de A.

Ejercicios

Ejercicios

5.1. Dada la señal discreta $h(n) = 1,\ n = -1, \ldots, 1$, obtener un filtro FIR $x(n),\ n = -3, \ldots, 3$ tal que $x * h - \delta$ tenga la menor energía posible (x es la aproximación FIR de siete coeficientes en el sentido de mínimos cuadrados al filtro inverso de h).

5.2. Hallar la fórmula de la seudoinversa (5.3) de A en términos de su SVD.

5.2. Solución de menor norma de ecuaciones indeterminadas

Ahora consideramos un sistema de ecuaciones lineales (5.1) en que A es una matriz rectangular $m \times n$ de rango máximo, pero ahora $m < n$. Ahora existirá un número infinito de soluciones, pues a cada solución podemos sumar un elemento del núcleo de A, de dimensión $n - n > 0$ y obtendremos otra solución. De hecho es fácil comprobar que el conjunto de soluciones es exactamente

$$\mathbf{x}_0 + \ker A = \{\mathbf{x}_0 + \mathbf{v},\ \mathbf{v} \in \ker A\},$$

donde $\mathbf{x}_0$ es una solución concreta.

Los subconjuntos de este tipo de un espacio vectorial, es decir, los que son de la forma $\{\mathbf{v}_0 + V\}$ con V un subespacio, se denominan *variedades afines*. El conjunto de soluciones es por tanto una variedad afín de $\mathbf{C}^n$. Nos planteamos ahora hallar entre todas estas la solución de menor norma.

El espacio de soluciones está dado por las ecuaciones

$$\mathbf{f}_i^*\mathbf{x} = b_i$$

donde los vectores fila $\mathbf{f}_i^*$ son las filas de A y los b_i a los coeficientes de $\mathbf{b}$.

Si $\mathbf{x}_0$ es una solución, las demás soluciones se obtienen sumando a $\mathbf{x}_0$ un vector ortogonal a los $\mathbf{f}_i$, es decir, un elemento del núcleo de A. De aquí se desprende que si $\mathbf{x}_0$ es ortogonal al núcleo, $\mathbf{x}_0$ es la solución de menor norma (pues cualquier otra solución $\mathbf{x}_0 + \mathbf{v}$, $\mathbf{v} \in \ker A$

tendrá norma al cuadrado $\|\mathbf{x}_0\|^2 + \|\mathbf{v}\|^2$). Así que $\mathbf{x}_0$ está en complemento ortogonal del núcleo de A:

$$\begin{aligned} \ker A = \{\mathbf{f}_1, \ldots, \mathbf{f}_m\}^\perp &= \text{span}\,\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}^\perp \\ \mathbf{x}_0 \in (\ker A)^\perp &= \text{span}\,\{\mathbf{f}_1, \ldots, \mathbf{f}_m\}^{\perp\perp} = \text{span}\,\{\mathbf{f}_1, \ldots, \mathbf{f}_m\} \end{aligned}$$

luego podemos escribir $\mathbf{x}_0$ como

$$\mathbf{x}_0 = \sum_{i=1}^{m} c_i \mathbf{f}_i = A^*\mathbf{c},\ \mathbf{c} = (c_1, \ldots, c_m)^\top.$$

Por tanto

$$AA^*\mathbf{c} = \mathbf{b} \Rightarrow \mathbf{c} = (AA^*)^{-1}\mathbf{b} \Rightarrow \mathbf{x}_0 = A^*(AA^*)^{-1}\mathbf{b}.$$

Para ver que AA^* es regular se puede utilizar la SVD de A. La matriz

$$A^+ = A^*(AA^*)^{-1} \tag{5.4}$$

es la seudoinversa de A.

Ejercicios

5.3. Hallar la fórmula de la seudoinversa (5.4) de A en términos de su SVD.

5.4. Realizar un programa que filtre paso bajo una imagen y luego obtenga una aproximación a la imagen original a partir de la filtrada.

El filtrado (convolución) se realizará por filas y la imagen resultante estará formada por aquellos píxeles de la imagen filtrada que sean combinación lineal de píxeles de la imagen original, sin que intervengan en su cálculo los ceros con que se extiende la imagen inicial para realizar la convolución (si la imagen tiene M columnas y el filtro tiene una longitud N, la imagen filtrada tendrá $M+N-1$ columnas). Este proceso se puede modelar como la multiplicación de cada fila de la imagen original por una matriz con filas linealmente independientes. Para la recuperación aproximada de la imagen original se obtendrá la imagen de menor energía que pueda dar lugar a la imagen filtrada.

5.3. Sistemas de ecuaciones generales

Dada la matriz A, $m \times n$, de rango $r \leq \text{mín}\{m, n\}$, y $\mathbf{b} \in \mathbf{C}^m$, buscamos los $\mathbf{x} \in \mathbf{C}^n$ que minimizan $\|A\mathbf{x} - \mathbf{b}\|$.

Aplicando la SVD a $A = U\Sigma V^*$,

$$\|A\mathbf{x} - \mathbf{b}\| = \|U\Sigma V^*\mathbf{x} - \mathbf{b}\| = \|\Sigma \underbrace{V^*\mathbf{x}}_{\mathbf{x}'} - \underbrace{U^*\mathbf{b}}_{\mathbf{b}'}\|,$$

vemos que podemos buscar los $\mathbf{x}'$ que minimizan

$$\|\Sigma\mathbf{x}' - \mathbf{b}'\|^2 = \sum_{k=1}^{r} |\sigma_k x'_k - b'_k|^2 + \sum_{k=r+1}^{m} |b'_k|^2,$$

que obviamente son los

$$\mathbf{x} \in \{(b'_1/\sigma_1, \ldots, b'_r/\sigma_r, x'_{r+1}, \ldots, x'_n), x'_{r+1}, \ldots, x'_n \in \mathbf{C}\}$$

es decir, los elementos de la variedad afín

$$\underbrace{(b'_1/\sigma_1, \ldots, b'_r/\sigma_r, \underbrace{0, \ldots, 0}_{(n-r)})}_{\mathbf{x}'_0} + \operatorname{span}\{\mathbf{e}_{r+1}, \ldots, \mathbf{e}_n\},$$

cuyo elemento de menor norma es $\mathbf{x}'_0 = \Sigma^+\mathbf{b}'$, si definimos Σ^+ como la matriz que se obtiene al transponer Σ y sustituir sus elementos no nulos por sus inversos. Esta solución corresponde a

$$\mathbf{x}_0 = V\mathbf{x}'_0 = \underbrace{V\Sigma^+U^*}_{A^+}\mathbf{b}.$$

La matriz A es la *seudoinversa* de A.

Ejercicios

5.5. Dar una fórmula de la seudoinversa de una matriz genérica utilizando la SVD de la matriz.

5.4. Mínimos cuadrados recursivos

Consideramos una secuencia de problemas como el de la sección 5.1 en la que sucesivamente se van añadiendo nuevas filas a la matriz A. Esta secuencia de problemas modela la situación en la que la información acerca del vector incógnita se va obteniendo de forma sucesiva y en cada instante se desea tener una aproximación de su valor correspondiente a los datos disponibles.

Nuestra secuencia de problemas es

$$A_k\mathbf{x}_k = \mathbf{b}_k,\ A_k = \begin{pmatrix} \mathbf{f}_1^* \\ \vdots \\ \mathbf{f}_k^* \end{pmatrix},\ \mathbf{b}_k = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}. \tag{5.5}$$

Sabemos que las soluciones en el sentido de mínimos cuadrados son

$$\mathbf{x}_k = (A_k^*A_k)^{-1}A_k^*\mathbf{b}_k,$$

pero el cálculo de las matrices $(A_k^*A_k)^{-1}$ se puede simplificar si conocemos su valor en las iteraciones anteriores. La clave está en la relación

$$A_k^*A_k = \sum_{r=1}^{k} \mathbf{f}_r\mathbf{f}_r^* \Rightarrow A_k^*A_k = A_{k-1}^*A_{k-1} + \mathbf{f}_k\mathbf{f}_k^*$$

y en el *lema de inversión* o *fórmula de Sherman-Morrison*

$$(B + \mathbf{x}\mathbf{y}^*)^{-1} = B^{-1} - \frac{B^{-1}\mathbf{x}\mathbf{y}^*B^{-1}}{1 + \mathbf{y}^*B^{-1}\mathbf{x}} \tag{5.6}$$

que se demuestra en [9] (y es válida siempre que B y $B + \mathbf{x}\mathbf{y}^*$ sean regulares, lo que ocurre en particular si B es regular y $\mathbf{x}$ y $\mathbf{y}$ son suficientemente pequeños), que nos permite calcular fácilmente $(A_{k+1}^*A_{k+1})^{-1}$ a partir de $(A_k^*A_k)^{-1}$.

Vamos a obtener la fórmula de $\mathbf{x}_{k+1}$ a partir de datos a partir de $\mathbf{x}_k$, $(A_k^*A_k)^{-1}$, y de los nuevos datos $\mathbf{f}_{k+1}$, b_{k+1}. Definiendo $P_k = (A_k^*A_k)^{-1}$, tenemos

$$\begin{aligned} \mathbf{x}_{k+1} &= (A_{k+1}^*A_{k+1})^{-1}A_{k+1}^*\mathbf{b}_{k+1} = (A_k^*A_k + \mathbf{f}_{k+1}\mathbf{f}_{k+1}^*)^{-1}\begin{pmatrix} A_k \\ \mathbf{f}_{k+1}^* \end{pmatrix}^*\begin{pmatrix} \mathbf{b}_k \\ b_{k+1} \end{pmatrix} \\ &= \underbrace{\left(P_k - \frac{P_k\mathbf{f}_{k+1}\mathbf{f}_{k+1}^*P_k}{1 + \mathbf{f}_{k+1}^*P_k\mathbf{f}_{k+1}}\right)}_{P_{k+1}}(A_k^*\mathbf{b}_k + b_{k+1}\mathbf{f}_{k+1}). \end{aligned} \tag{5.7}$$

Definiendo el *vector de Kalman*

$$\mathbf{k}_{k+1} = \frac{P_k\mathbf{f}_{k+1}}{1 + \mathbf{f}_{k+1}^*P_k\mathbf{f}_{k+1}} \tag{5.8}$$

queda

$$\mathbf{x}_{k+1} = \underbrace{\mathbf{x}_k - \mathbf{k}_{k+1}\mathbf{f}_{k+1}^*\mathbf{x}_k}_{P_{k+1}A_k^*\mathbf{b}_k} + b_{k+1}P_{k+1}\mathbf{f}_{k+1}. \tag{5.9}$$

Por otra parte, usando la fórmula de recursión de P_{k+1} (en (5.7)) obtenemos la relación

$$P_{k+1}\mathbf{f}_{k+1} = \mathbf{k}_{k+1},$$

que utilizada en (5.9) nos lleva a

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{k}_{k+1}(b_{k+1} - \mathbf{f}_{k+1}^*\mathbf{x}_k). \tag{5.10}$$

El siguiente algoritmo permite calcular eficazmente la solución de (5.5) para $k+1$ utilizando el resultado $\mathbf{x}_k$ y la matriz P_k. Se puede esperar a tener una matriz A_k cuadrada para comenzar, o utilizar la inicialización heurística $\mathbf{x}_0 = \mathbf{0}$, $P_0 = \delta^{-1}I$, δ pequeño.
En cada etapa, con los nuevos datos b_{k+1}, $\mathbf{f}_{k+1}$,
(1) calcular $\mathbf{k}_{k+1}$ usando (5.8),
(2) calcular $\mathbf{x}_{k+1}$ usando (5.10),
(3) calcular $P_{k+1} = P_k - \mathbf{k}_{k+1}\mathbf{f}_{k+1}^*P_k$.

Ejercicios

5.6. En este ejercicio se considera la implementación de un igualador adaptativo (ver figura 5.1). Implementar el algoritmo de mínimos cuadrados recursivos para el cálculo del filtro FIR $h(n)$ ($n = 0, \ldots, m-1$) que cuando procese la señal de entrada $f(n)$ ($n \geq 0$) genere una salida que aproxime de la mejor forma posible la señal deseada $y(n)$ ($n \geq 0$). En este caso tendremos una nueva ecuación en los coeficientes $h(n)$ por cada nuevo par de datos $(f(n), y(n))$, $n \geq m-1$.

De acuerdo con la fórmula de la convolución $h(n) = \sum_{r=0}^{m-1} h(r)f(n-r)$ la primera ecuación será

$$y(m-1) = h(0)f(m-1) + \ldots + h(m-1)f(0).$$

Aplicarlo al caso en el que la señal recibida es $f = z * g + e$, $g(n) = 1/3$, $n = 0, 1, 2$, $z(n)$ es una señal aleatoria de muestras gaussianas independientes de media nula y varianza unidad, $e(n)$ es otra señal del mismo tipo, independiente de la anterior y de varianza 0.1, y queremos recuperar una versión retardada de $z(n)$ mediante el filtro $h(n)$, $n = 0, \ldots, 9$.

Aplicar el programa para el cálculo de una aproximación de longitud $m = 15$ al filtro inverso del filtro $g(n) = 1/3$, $n = 1, \ldots, 3$. Se utilizará una señal de ruido blanco $r(n)$ que será procesada por el filtro $g(n)$ dando lugar a la señal $f(n)$. La señal $y(n)$ será una versión retrasada de la señal $r(n)$.

5.5. Mínimos cuadrados totales

Partimos de nuevo de la ecuación

$$A\mathbf{x} = \mathbf{b}$$

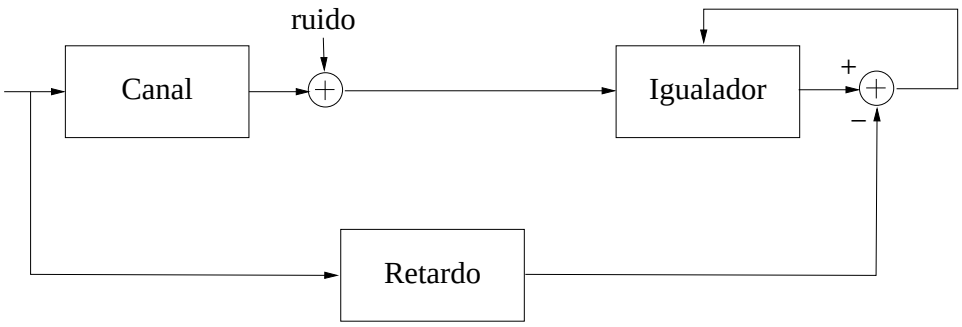

Figura 5.1: Esquema de igualador adaptativo.

en la que A, es $m \times n$, $m > n$, de rango máximo. Suponemos que no tiene solución, es decir, $\mathbf{b}$ no pertenece a la imagen de A. El problema de *mínimos cuadrados totales* consiste en hallar la perturbación de menor norma aplicada tanto a A como a $\mathbf{b}$ que haga que el sistema tenga solución. Con más precisión, la ecuación puede escribirse como

$$\begin{pmatrix} A & \mathbf{b} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ -1 \end{pmatrix} = \mathbf{0}.$$

La ecuación perturbada es

$$\left(\begin{pmatrix} A & \mathbf{b} \end{pmatrix} + \begin{pmatrix} E & \mathbf{w} \end{pmatrix} \right) \begin{pmatrix} \mathbf{x} \\ -1 \end{pmatrix} = \mathbf{0}.$$

Queremos hallar la matriz de perturbación de menor norma que haga que el sistema tenga solución, y obtener la solución corresponiente. Se puede demostrar que la matriz perturbada que buscamos es la que se obtiene de la SVD de $\begin{pmatrix} A & \mathbf{b} \end{pmatrix}$ anulando su menor valor singular, σ_n. El vector buscado $\begin{pmatrix} \mathbf{x} \\ -1 \end{pmatrix}$ estará en el núcleo de la matriz perturbada, siendo por tanto proporcional al vector singular por la derecha $\mathbf{v}_n$ correspondiente al valor singular anulado.

Ejercicios

5.7. Consideramos el problema de ajustar una recta que pase por el origen a un conjunto de puntos observados (x_i, y_i). Comprobar que

- Si aplicamos mínimos cuadrados a la ecuación $\mathbf{x}a = \mathbf{y}$, donde $\mathbf{x} = (x_1, \ldots, x_m)^T$, $\mathbf{y} = (y_1, \ldots, y_m)^T$, y a es la incógnita, obtenemos la solución que minimiza la suma de las distancias en vertical al cuadrado de cada punto a la recta. ¿Qué tendríamos que hacer para minimizar la suma de las distancias en horizontal al cuadrado?
- Si aplicamos PCA para minimizar $\sum_{i=1}^{m} (\alpha x_i + \beta y_i)^2$ con la restricción $\alpha^2 + \beta^2 = 1$, estamos minimizando la suma de las distancias al cuadrado a la recta. Concluir que la solución (α, β) que

obtenemos es un vector unitario del núcleo de la matriz de rango inferior más cercana a $(\mathbf{x}\ \ \mathbf{y})$ en la norma de Frobenius (definida en la sección 4.2).

- Si aplicamos mínimos cuadrados totales, obtenemos la misma recta que en el caso anterior. Este resultado se entiende mejor comprobando que para una matriz de rango uno la norma inducida y la norma de Frobenius coinciden.

5.6. Regresión ortogonal

Ahora consideramos el problema de la obtención de la recta en el plano euclídeo que minimiza la suma de las distancias al cuadrado de una serie de puntos $\mathbf{p}_i$. Utilizando coordenadas homogéneas notamos estos puntos como $\mathbf{p}_i = (x_i, y_i, 1)$ y la recta como $l_1x + l_2y + l_3 = 0$, es decir, de coordenadas homogéneas $\mathbf{l} = (l_1, l_2, l_3)$, normalizadas de forma que $l_1^2 + l_2^2 = 1$. Entonces la distancia de $\mathbf{p}_i$ a la recta se puede escribir como

$$d_i = \frac{|l_1x_i + l_2y_i + l_3|}{\sqrt{l_1^2 + l_2^2}} = |\mathbf{l}^\top \mathbf{p}_i|.$$

Por tanto queremos minimizar

$$\sum_i \mathbf{l}^\top \mathbf{p}_i \mathbf{p}_i^\top \mathbf{l} = \mathbf{l}^\top \underbrace{\left(\sum_i \mathbf{p}_i \mathbf{p}_i^\top \right)}_{E} \mathbf{l}$$

con la restricción

$$1 = l_1^2 + l_2^2 = \mathbf{l}^\top \underbrace{\begin{pmatrix} 1 & & \\ & 1 & \\ & & 0 \end{pmatrix}}_{J} \mathbf{l}.$$

Por el teorema de los multiplicadores de Lagrange sabemos que los posibles extremos condicionados verificarán

$$\nabla \left(\mathbf{l}^\top E \mathbf{l} - \lambda \mathbf{l}^\top J \mathbf{l} \right) = 2(E - \lambda J)\mathbf{l} = 0.$$

Un vector $\mathbf{l}$ que verifique esta condición tendrá un coste $C = \mathbf{l}^\top E \mathbf{l}$ que podemos calcular por la relación

$$\mathbf{l}^\top \underbrace{(E - \lambda J)\mathbf{l}}_{0} = C - \lambda \Rightarrow C = \lambda.$$

Por tanto las coordenadas de la recta que buscamos están en el núcleo por la derecha de $E - \lambda J$ donde λ es la menor raíz de la ecuación $|E - \lambda J| = 0$.

Ejercicios

5.8. Modificar la técnica de regresión ortogonal para obtener la mejor aproximación por un plano de un conjunto de puntos en el espacio euclídeo tridimensional.

5.7. Mínimos cuadrados y estimación estadística. Mínimos cuadrados robustos

La técnica de mínimos cuadrados se justifica desde un punto de vista probabilístico como la estimación de máxima verosimilitud de un vector $\mathbf{x}$ a partir de observaciones escalares b_i obtenidas como

$$b_i = \mathbf{a}_i^* \mathbf{x} + e_i$$

donde las e_i son variables aleatorias gaussianas independientes de media nula y la misma desviación típica σ.

En efecto, como la función de densidad de probabilidad del vector $\mathbf{e} = (e_1, \ldots, e_m)$ es

$$f(\mathbf{e}) = \frac{1}{(2\pi)^{m/2}\sigma^m} e^{-\frac{\|\mathbf{e}\|^2}{2\sigma^2}}.$$

el vector $\mathbf{e}$ más verosímil será el de menor normal, y corresponderá con el valor $\mathbf{x}$ que minimice

$$\sum_{i=1}^{m} |e_i|^2 = \sum^{m} |b_i - \mathbf{a}_i^* \mathbf{x}|^2 = \|\mathbf{b} - A\mathbf{x}\|^2$$

donde A es la matriz cuyas filas son las $\mathbf{a}_i^*$.

En la práctica es frecuente que una cierta proporción de los datos *(outliers)* no esté generado de acuerdo con este modelo, sino que esté afectado de otras perturbaciones con mayor varianza. La técnica de *RAndom SAmple Consensus (RANSAC)* está pensada para tratar este problema [2, p. 116]. La idea básica consiste en aplicar mínimos cuadrados a un subconjunto de los datos que esté libre de outliers. Para conseguirlo hacemos un cierto número de estimaciones de $\mathbf{x}$ con el número mínimo posible de muestras, es decir, la dimensión de $\mathbf{x}$, n.

Suponiendo que disponemos de m observaciones, el número N de subconjuntos de tamaño m que tenemos que tomar para que la probabilidad de que al menos uno esté libre de outliers sea p (se suele tomar $p = 0{,}99$) es

$$N = \frac{\log(1-p)}{\log[1-(1-\epsilon)^m]}. \tag{5.11}$$

El algoritmo que presentamos supone que conocemos a priori la desviación típica σ de las observaciones que se ajustan al modelo gaussiano *(inliers)* pero no la proporción ϵ de outliers. Llamamos t al valor tal que la probabilidad de que $|e_i|$ sea menor que t sea α (típicamente $\alpha = 0{,}95$). Dado un valor de $\mathbf{x}$ suficientemente cercano al verdadero, el número de observaciones tales que $|\mathbf{a}_i^* \mathbf{x} - b_i| < t$ nos da una estimación del número

de *inliers* de la muestra, suponiendo que los *outliers* están con gran probabilidad más alejados.

El algoritmo es el siguiente. La idea es estimar en cada iteración el parámetro ϵ utilizando el umbral t y parar cuando el número de iteraciones supere N para este ϵ dado por (5.11).

- Inicializar `N`$= \infty$, `numero_iteraciones`$= 0$.
- Mientras `N`$>$`numero_iteraciones`,
  - Elegir al azar una muestra de m observaciones.
  - Estimar el número de *inliers* usando el umbral t.
  - Estimar la proporción de *outliers* $\epsilon = 1-$(número de *inliers*)/(número de puntos).
  - Recalcular `N` $=$ mín$\{$`N`$, N(\epsilon)\}$, donde $N(\epsilon)$ está dado por (5.11).
  - Incrementar `numero_iteraciones`.
- Aplicar mínimos cuadrados a los *inliers* del subconjunto para el que se ha encontrado un mayor número de ellos.

Ejercicios

5.9. Demostrar la fórmula (5.11).

5.10. Generar un conjunto de 100 datos (a_i, b_i) donde los a_i tienen distribución uniforme entre 0 y 1 y $b_i = a_i x_1 + x_2 + e_i$ con probabilidad 0,95 y $b_i = a_i x_1 + x_2 + 10 e_i$ con probabilidad 0,05, siendo $x_1 = x_2 = 1$ y las e_i variables gaussianas de media nula y desviación típica 0.1. Estimar (x_1, x_2) a partir de los pares (a_i, b_i) mediante mínimos cuadrados y mediante el algoritmo RANSAC.

Capítulo 6

Condicionamiento de un problema

6.1. Número de condición de un problema

El *número de condición absoluto* de un problema nos informa de cómo varía el resultado $f(\mathbf{x})$ del problema si se perturban un poco los datos $\mathbf{x}$. Se define como

$$\text{cond abs}_{\mathbf{x}} f = \lim_{\epsilon \to 0} \sup_{\|\delta\mathbf{x}\| \leq \epsilon} \frac{\|f(\mathbf{x}+\delta\mathbf{x}) - f(\mathbf{x})\|}{\|\delta\mathbf{x}\|}. \tag{6.1}$$

Si la función es diferenciable en $\mathbf{x}$ se demuestra que el límite es la norma de la matriz jacobiana de f, $\|Df_{\mathbf{x}}\|$ (ver apéndice 6.7).

Vamos a obtener una expresión que nos permite en algunos casos importantes calcular el número de condición absoluto sin necesidad de derivar una expresión explícita de f (aunque no la usaremos hasta 6.3). Definiendo $\mathbf{y}(t) = f(\mathbf{x}(t))$, por la regla de la cadena,

$$\mathbf{y}'(0) = Df_{\mathbf{x}(0)}\mathbf{x}'(0).$$

Si $\mathbf{x}(t) = \mathbf{x} + t\mathbf{e}$, $\mathbf{y}'(0) = Df_{\mathbf{x}}\mathbf{e}$, y por tanto

$$\begin{aligned} \|Df_{\mathbf{x}}\| = \sup_{\|\mathbf{e}\|=1} \|Df_{\mathbf{x}}\mathbf{e}\| = \sup_{\|\mathbf{e}\|=1} \|\mathbf{y}'(0)\| \\ \Rightarrow \text{ cond abs}_{\mathbf{x}} f = \|D_{\mathbf{x}} f\| = \sup_{\|\mathbf{e}\|=1} \|\mathbf{y}'(0)\|. \end{aligned} \tag{6.2}$$

El interés de esta fórmula radica en que $\mathbf{y}'(0)$ se puede obtener a veces sin necesidad de la expresión explícita de f.

El *número de condición relativo* se define de forma análoga, pero mediante un cociente de errores relativos:

$$\mathrm{cond}_{\mathbf{x}} f = \lim_{\epsilon \to 0} \sup_{\|\delta \mathbf{x}\| \leq \epsilon} \frac{\dfrac{\|f(\mathbf{x}+\delta\mathbf{x}) - f(\mathbf{x})\|}{\|f(\mathbf{x})\|}}{\dfrac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|}} = \frac{\|\mathbf{x}\|}{\|f(\mathbf{x})\|} \mathrm{cond\ abs}_{\mathbf{x}} f. \tag{6.3}$$

6.2. Número de condición de una matriz regular

El *número de condición* $\kappa(A)$ de una matriz A con núcleo trivial (es decir, núcleo $\{\mathbf{0}\}$, o sea, con columnas linealmente independientes, cuadrada o no) nos da una idea de cómo puede variar en términos relativos, como máximo, el producto $A\mathbf{x}$ cuando $\mathbf{x}$ sufre una perturbación. Corresponde al número de condición relativo del problema de multiplicar la matriz por un vector $\mathbf{x}$ ($f(\mathbf{x}) = A\mathbf{x}$), tomando el caso peor como valor de este vector:

$$\mathrm{cond}_{\mathbf{x}} f = \lim_{\epsilon \to 0} \sup_{\|\delta \mathbf{x}\| \leq \epsilon} \frac{\dfrac{\|A\delta\mathbf{x}\|}{\|A\mathbf{x}\|}}{\dfrac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|}} = \frac{\|\mathbf{x}\|}{\|A\mathbf{x}\|} \sup_{\|\delta \mathbf{x}\| \leq \epsilon} \frac{\|A\delta\mathbf{x}\|}{\|\delta\mathbf{x}\|} = \frac{\|\mathbf{x}\|}{\|A\mathbf{x}\|}\sigma_1$$

$$\kappa(A) = \sup_{\mathbf{x}} \mathrm{cond}_{\mathbf{x}} f = \sigma_1 \sup_{\mathbf{x}} \frac{\|\mathbf{x}\|}{\|A\mathbf{x}\|} = \frac{\sigma_1}{\inf\limits_{\mathbf{x}} \dfrac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}} = \frac{\sigma_1}{\sigma_n} = \|A\|\|A^{-1}\|.$$

Obsérvese que si intentásemos aplicar esta definición a matrices con núcleo no trivial (es decir, con columnas linealmente dependientes, como por ejemplo cualquier matriz $m \times n$, $m < n$), el denominador $\inf_{\mathbf{x}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$ valdría cero, con lo que el número de condición de la matriz sería $+\infty$ si A no es la matriz nula, y no estaría definido si A es la matriz nula. En adelante, salvo que lo indiquemos expresamente, consideraremos solamente matrices con columnas linealmente independientes.

Dos propiedades inmediatas son
(1) $\kappa(A) = \dfrac{\sigma_1}{\sigma_n} \geq 1$,
(2) $\kappa(A) = \kappa(A^{-1})$.

El inverso del número de condición se puede interpretar como la distancia d de la matriz al conjunto de las matrices de rango inferior normalizada por la norma (valga la redundancia) de la matriz:

$$\frac{1}{\kappa(A)} = \frac{\sigma_n}{\sigma_1} = \frac{d}{\|A\|}.$$

El número de condición de la matriz acota superiormente el número de condición relativo de la multiplicación de una matriz por un vector respecto de perturbaciones de la matriz, es decir, del problema $f(A) = A\mathbf{b}$. En efecto,

$$\mathrm{cond}_A f = \sup_{\delta A} \frac{\frac{\|(\delta A)\mathbf{b}\|}{\|A\mathbf{b}\|}}{\frac{\|\delta A\|}{\|A\|}} = \frac{\|A\|}{\|A\mathbf{b}\|} \sup_{\delta A} \frac{\|(\delta A)\mathbf{b}\|}{\|\delta A\|} = \frac{\|A\|\|\mathbf{b}\|}{\|A\mathbf{b}\|} \leq \frac{\sigma_1}{\sigma_n}. \tag{6.4}$$

6.3. Condicionamiento de la resolución de un sistema lineal

El número de condición de una matriz regular es también el número de condición relativo del problema de hallar $\mathbf{y}$ tal que $A\mathbf{y} = \mathbf{b}$ cuando $\mathbf{b}$ es exacto pero A puede estar perturbada (es decir, $f(A) = A^{-1}\mathbf{b}$).

Podemos obtener este resultado utilizando la fórmula (6.2). En nuestro caso,

$$(A + tE)\mathbf{y}(t) = \mathbf{b},\ A\mathbf{y}(0) = \mathbf{b},$$

Derivando respecto de t y notando $\mathbf{y} \equiv \mathbf{y}(0)$,

$$E\mathbf{y}(t) + A\mathbf{y}'(t) = 0 \Rightarrow E\mathbf{y}(0) + A\mathbf{y}'(0) = 0 \Rightarrow \mathbf{y}'(0) = -A^{-1}E\mathbf{y}.$$

Utilizando la primera de las siguientes relaciones, fáciles de demostrar,

$$\sup_{\|E\|=1} \|AE\mathbf{x}\| = \|A\|\|\mathbf{x}\|,$$
$$\sup_{\|E\|=1} \|AEB\| = \|A\|\|B\|,$$

tenemos

$$\sup_{\|E\|=1} \|\mathbf{y}'(0)\| = \sup_{\|E\|=1} \|A^{-1}E\mathbf{y}\| = \|A^{-1}\|\|\mathbf{y}\|$$
$$\Rightarrow \mathrm{cond}_A f = \frac{\|A\|}{\|\mathbf{y}\|}\|A^{-1}\|\|\mathbf{y}\| = \kappa(A).$$

Ejercicios

6.1. Calcular el número de condición de la matriz de desconvolución del ejercicio 5.4.

6.4. Condicionamiento del problema de mínimos cuadrados

Volvemos al problema planteado en la sección 5.1 para estudiar su condicionamiento. En esta sección seguimos básicamente [1].

Estudiamos primero la sensibilidad del problema respecto de perturbaciones en el vector $\mathbf{b}$ ($f(\mathbf{b}) = (A^*A)^{-1}A^*\mathbf{b} = \mathbf{y}$).

Conocemos el número de condición de este problema (ecuación (6.4)). Si θ es el ángulo que forma $\mathbf{b}$ con su proyección ortogonal sobre el rango de A, $A\mathbf{y}$ (ver figura 6.4), tenemos que $\|\mathbf{b}\| = \|A\mathbf{y}\|/\cos\theta$, luego

$$\begin{aligned}\operatorname{cond}_{\mathbf{b}} f &= \frac{\|(A^*A)^{-1}A^*\|\|\mathbf{b}\|}{\|\mathbf{y}\|} = \frac{\|(A^*A)^{-1}A^*\|\|A\mathbf{y}\|}{\|\mathbf{y}\|\cos\theta} \\ &= \underbrace{\|(A^*A)^{-1}A^*\|}_{\sigma_n^{-1}}\underbrace{\|A\|}_{\sigma_1}\frac{1}{\cos\theta}\frac{\|A\mathbf{y}\|}{\|A\|\|\mathbf{y}\|} = \frac{\kappa(A)}{\cos\theta}\frac{\|A\mathbf{y}\|}{\|A\|\|\mathbf{y}\|} \\ &\leq \frac{\kappa(A)}{\cos\theta}.\end{aligned}$$

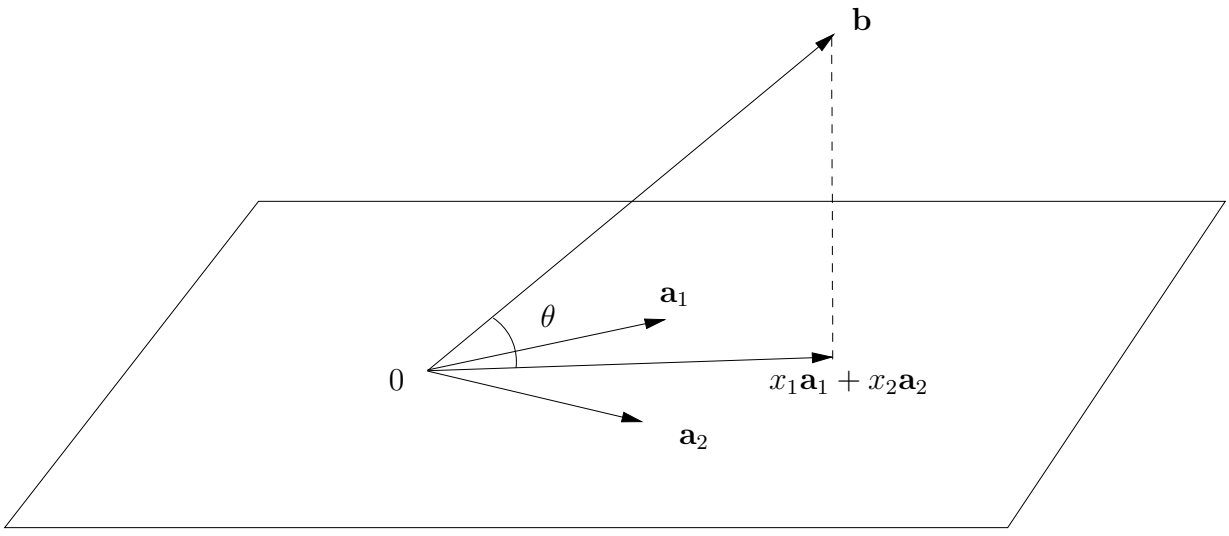

Figura 6.1: Elementos del problema de mínimos cuadrados para $A = (\mathbf{a}_1, \mathbf{a}_2)$, $\mathbf{y} = (x_1, x_2)^\top$.

Veamos ahora cómo se modifica la solución $\mathbf{y}$ si lo que se perturba ligeramente es la matriz A ($f(A) = (A^*A)^{-1}A^*\mathbf{b} = \mathbf{y}$). Si añadimos $\delta A = tE$ a la matriz A, con $\|E\| = 1$,

la solución del problema $\mathbf{y}(t)$ verifica

$$(A+tE)^\top (A+tE)\mathbf{y}(t) = (A+tE)^\top \mathbf{b}.$$

Derivando respecto de t y particularizando en $t=0$ nos queda

$$E^\top A\mathbf{y} + A^\top E\mathbf{y} + A^\top A\mathbf{y}'(0) = E^\top \mathbf{b} \Rightarrow \mathbf{y}'(0) = -(A^*A)^{-1}A^*E\mathbf{y} + (A^*A)^{-1}E^\top \mathbf{r},$$

donde

$$\mathbf{r} = \mathbf{b} - A\mathbf{y}$$

es el error de la proyección ortogonal de $\mathbf{b}$ sobre la imagen de A. Tenemos

$$\mathrm{cond}_A f = \frac{\|A\|}{\|\mathbf{y}\|} \sup_{\|E\|=1} \|\mathbf{y}'(0)\| \leq \|(A^*A)^{-1}A^*\|\|A\| + \frac{\|(A^*A)^{-1}\|\|A\|\|\mathbf{r}\|}{\|\mathbf{y}\|}.$$

El primer sumando vale $\kappa(A)$. Para el segundo sumando observamos que $\|\mathbf{r}\| = \|A\mathbf{y}\| \tan\theta$, luego y

$$\begin{aligned}\frac{\|(A^*A)^{-1}\|\|A\|\|\mathbf{r}\|}{\|\mathbf{y}\|} &= \|(A^*A)^{-1}\|\|A\| \frac{\|A\mathbf{y}\| \tan\theta}{\|\mathbf{y}\|} \\ &= \kappa(A)^2 \tan\theta \frac{\|A\mathbf{y}\|}{\|A\|\|\mathbf{y}\|}\end{aligned}$$

y tenemos

$$\mathrm{cond} = \kappa(A) + \kappa(A)^2 \tan\theta \frac{\|A\mathbf{y}\|}{\|A\|\|\mathbf{y}\|} \leq \kappa(A) + \kappa(A)^2 \tan\theta.$$

6.5. Condicionamiento de la solución de menor norma de sistemas indeterminados

Consideramos el problema de la sección 5.2, en el que queremos hallar el vector $\mathbf{y}$ de menor norma que es solución del sistema $A\mathbf{y} = \mathbf{b}$, con A $m \times n$, $m < n$, de rango m. S

Sabemos que $\mathbf{y} = A^*(AA^*)^{-1}$, luego el condicionamiento respecto de perturbaciones en $\mathbf{b}$ (problema $f(\mathbf{b}) = \mathbf{y}$) será el del producto de la matriz $A^*(AA^*)^{-1}$ por $\mathbf{b}$, que por la sección 6.2 vale

$$\mathrm{cond} f = \frac{\|\mathbf{y}\|}{\|\mathbf{b}\|}\sigma_1(A^*(AA^*)^{-1}) = \frac{\|\mathbf{y}\|}{\|\mathbf{b}\|}\sigma_m^{-1}(A)$$

y está acotado por

$$\kappa(A^*(AA^*)^{-1}) = \frac{\sigma_1(A^*(AA^*)^{-1})}{\sigma_n(A^*(AA^*)^{-1})} = \frac{\sigma_n^{-1}(A)}{\sigma_1^{-1}(A)} = \frac{\sigma_1(A)}{\sigma_n(A)} = \kappa(A).$$

Para el condicionamieno respecto de A partimos de las ecuaciones

$$\mathbf{y}(t) = (A`tE)^*\mathbf{c}(t), \;\; (A+tE)(A+tE)^*\mathbf{c}(t),$$

con E unitaria, que derivamos respecto de t y particularizamos en $t=0$. Después de eliminar $\mathbf{c}(0)$ y $\mathbf{c}'(0)$ obtenemos (ordenando los términos como en [1])

$$\mathbf{y}'(0) = [I - A^*(AA^*)^{-1}A]E^*(AA^*)^{-1}\mathbf{b} - A^*(AA^*)^{-1}E\mathbf{y}.$$

Utilizando

$$\|I - A^*(AA^*)^{-1}A\| = 1$$

y

$$\|\mathbf{y}\| = \|A^*(AA^*)^{-1}\mathbf{b}\| \geq \sigma_m(A)\|(AA^*)^{-1}\mathbf{b}\| \Rightarrow \|(AA^*)^{-1}\mathbf{b}\| \leq \frac{\|\mathbf{y}\|}{\sigma_m(A)}$$

tenemos

$$\begin{aligned}\|\mathbf{y}'(0)\| &\leq \|[I - A^*(AA^*)^{-1}A]E^*(AA^*)^{-1}\mathbf{b}\| + \|A^*(AA^*)^{-1}E\mathbf{y}\| \\ &\leq \|I - A^*(AA^*)^{-1}A\|\|E^*\|\|(AA^*)^{-1}\mathbf{b}\| + \|A^*(AA^*)^{-1}\|\|E\|\|\mathbf{y}\| \\ &\leq \frac{\|\mathbf{y}\|}{\sigma_m(A)} + \frac{\|\mathbf{y}\|}{\sigma_m(A)} = 2\frac{\|\mathbf{y}\|}{\sigma_m(A)}\end{aligned}$$

luego

$$\mathrm{cond}f = \frac{\|A\|}{\|\mathbf{y}\|}\|\mathbf{y}'(0)\| \leq 2\frac{\sigma_1(A)}{\sigma_m(A)} = 2\kappa(A^*).$$

Obsérvese que no aparecen términos en κ^2 como en el problema estándar de mínimos cuadrados.

6.6. Condicionamiento del problema de autovalores de matrices hermíticas

Queremos obtener el número de condición del problema que asigna a una matriz uno de sus autovectores (suponemos que todos los autovectores son distintos, que es el caso genérico). Para ello partimos, en la línea de [1, 7.2.2], de

$$A\mathbf{y} = \lambda\mathbf{y}, \; \mathbf{y}^*\mathbf{y} = 0,$$

y tomamos una matriz simétrica E de norma unidad y consideramos las ecuaciones

$$(A+tE)\mathbf{y}(t) = \lambda(t)\mathbf{y}(t), \;\; \mathbf{y}(t)^*\mathbf{y}(t) = 1$$

que derivando y tomando $t = 0$, con $\mathbf{y} \equiv \mathbf{y}(0)$, nos dan

$$E\mathbf{y} + A\mathbf{y}'(0) = \lambda'(0)\mathbf{y} + \lambda\mathbf{y}'(0), \;\; \mathbf{y}'(0)^*\mathbf{y} = 0.$$

Multiplicando la primera ecuación por $\mathbf{y}^*$ obtenemos (utilizando que A es hermítica)

$$\mathbf{y}^* E\mathbf{y} = \lambda'(0)$$

que nos proporciona el número de condición absoluto del problema $f(A) = \lambda$,

$$\text{cond abs}_A f = \sup_{\|E\|=1} |\lambda'(0)| = 1.$$

Por otra parte tenemos

$$(A - \lambda I)\mathbf{y}'(0) = (\mathbf{y}^* E\mathbf{y})\mathbf{y} - E\mathbf{y}$$

luego

$$\|(A - \lambda I)\mathbf{y}'(0)\| \leq 2\|\mathbf{y}\| = 2.$$

Si A es tiene autovalores λ_i, $A - \lambda I$ tiene autovalores $\lambda_i - \lambda$ y los mismos autovectores que A. Si λ tiene multiplicidad uno, la menor amplificación de la norma de un vector ortogonal a $\mathbf{y}$ al multiplicarlo por $A - \lambda I$ es

$$\min_{\lambda_i \neq \lambda} |\lambda_i - \lambda|,$$

con lo que

$$\|(A - \lambda I)\mathbf{y}'(0)\| \geq \min_{\lambda_i \neq \lambda} |\lambda_i - \lambda| \|\mathbf{y}'(0)\|$$

luego el número de condición del problema $f(A) = \mathbf{y}$ verifica

$$\text{cond abs}_A f = \sup_{\|E\|=1} \|\mathbf{y}'(0)\| \leq \frac{2}{\min_{\lambda_i \neq \lambda} |\lambda_i - \lambda|}.$$

6.7. Apéndice

Veamos que si f es diferenciable, su número de condición absoluto en $\mathbf{x}$ es

$$\lim_{\epsilon \to 0} \sup_{\|\delta\mathbf{x}\| \leq \epsilon} \frac{\|f(\mathbf{x} + \delta\mathbf{x}) - f(\mathbf{x})\|}{\|\delta\mathbf{x}\|} = \|Df_{\mathbf{x}}\|.$$

Definiendo $g(\delta\mathbf{x}) = \dfrac{\|f(\mathbf{x} + \delta\mathbf{x}) - f(\mathbf{x}) - Df_{\mathbf{x}}(\delta\mathbf{x})\|}{\|\delta\mathbf{x}\|}$, la diferenciabilidad de f en $\mathbf{x}$ equivale a que $\lim_{\delta\mathbf{x} \to 0} g(\delta\mathbf{x}) = 0$, y por lo tanto $\lim_{\epsilon \to 0} \sup_{\|\delta\mathbf{x}\| < \epsilon} g(\delta\mathbf{x}) = 0$. Tenemos

$$\frac{\|f(\mathbf{x} + \delta\mathbf{x}) - f(\mathbf{x})\|}{\|\delta\mathbf{x}\|} = \frac{\|Df_{\mathbf{x}}(\delta\mathbf{x}) + f(\mathbf{x} + \delta\mathbf{x}) - f(\mathbf{x}) - Df_{\mathbf{x}}(\delta\mathbf{x})\|}{\|\delta\mathbf{x}\|} \leq \frac{\|Df_{\mathbf{x}}(\delta\mathbf{x})\|}{\|\delta\mathbf{x}\|} + g(\delta\mathbf{x})$$

Por tanto

$$\sup_{\|\delta\mathbf{x}\|<\epsilon} \frac{\|f(\mathbf{x}+\delta\mathbf{x})-f(\mathbf{x})\|}{\|\delta\mathbf{x}\|} \leq \sup_{\|\delta\mathbf{x}\|<\epsilon} \frac{\|Df_{\mathbf{x}}(\delta\mathbf{x})\|}{\|\delta\mathbf{x}\|} = \|Df_{\mathbf{x}}\|.$$

Por otra parte,

$$\frac{\|f(\mathbf{x}+\delta\mathbf{x})-f(\mathbf{x})\|}{\|\delta\mathbf{x}\|} \geq \frac{\|Df_{\mathbf{x}}(\delta\mathbf{x})\|}{\|\delta\mathbf{x}\|} - g(\delta\mathbf{x})$$

luego

$$\sup_{\|\delta\mathbf{x}\|<\epsilon} \frac{\|f(\mathbf{x}+\delta\mathbf{x})-f(\mathbf{x})\|}{\|\delta\mathbf{x}\|} \geq \sup_{\|\delta\mathbf{x}\|<\epsilon} \frac{\|Df_{\mathbf{x}}(\delta\mathbf{x})\|}{\|\delta\mathbf{x}\|} - \sup_{\|\delta\mathbf{x}\|<\epsilon} g(\delta\mathbf{x})$$

y tomando límites,

$$\lim_{\epsilon\to 0} \sup_{\|\delta\mathbf{x}\|<\epsilon} \frac{\|f(\mathbf{x}+\delta\mathbf{x})-f(\mathbf{x})\|}{\|\delta\mathbf{x}\|} \geq \|Df_{\mathbf{x}}\| - \lim_{\epsilon\to 0} \sup_{\|\delta\mathbf{x}\|<\epsilon} g(\delta\mathbf{x}) = \|Df_{\mathbf{x}}\|.$$

Ejercicios

6.2. Comprobar que $\mathrm{cond}(A) = \mathrm{cond}(A^+)$.

6.3. Comprobar que $\|(A^*A)^{-1}\|\|A\|^2 = \kappa(A)^2$ y que $\|(A^*A)^{-1}A^*\|\|A\| = \kappa(A)$.

6.4. Poner un ejemplo de un problema de mínimos cuadrados $A\mathbf{x} = \mathbf{b}$ con A matriz 3×2 con valores altos de $\kappa(A)$ y θ. Estudiar experimentalmente qué ratios máximos encontramos entre errores relativos del resultado y errores relativos de perturbaciones, primero en A y lueog en $\mathbf{b}$.

Capítulo 7

Factorización QR

7.1. Motivación

En este capítulo vemos cómo toda matriz regular se puede factorizar mediante algoritmos computacionalmente eficientes como

$$A = QR$$

donde Q es unitaria y R triangular superior. Una aplicación evidente es la resolución de ecuaciones lineales, puesto que $A\mathbf{x} = \mathbf{b}$ se transforma en

$$R\mathbf{x} = Q^*\mathbf{b},$$

que se puede resolver despejando sucesivamente $x_n, x_{n-1}, \ldots, x_1$ *(retrosustitución)* (sin calcular R^{-1}).

La factorización QR es también permite la solución de problemas de mínimos cuadrados (sección 5.1). En efecto,

$$\begin{aligned}\|A\mathbf{x} - \mathbf{b}\|^2 = \|QS\mathbf{x} - \mathbf{b}\|^2 = \|S\mathbf{x} - \underbrace{Q^*\mathbf{b}}_{\hat{\mathbf{b}}}\|^2 &= \| \begin{pmatrix} R \\ 0_{m-n,n} \end{pmatrix} \mathbf{x} - \begin{pmatrix} \hat{\mathbf{b}}_{1,\ldots,n} \\ \hat{\mathbf{b}}_{n+1,\ldots,m} \end{pmatrix} \|^2 \\ &= \|R\mathbf{x} - \hat{\mathbf{b}}_{1,\ldots,n}\|^2 + \|\hat{\mathbf{b}}_{n+1,\ldots,m}\|^2.\end{aligned}$$

que se minimiza obviamente con $\mathbf{x} = R^{-1}\hat{\mathbf{b}}_{1,\ldots,n}$. Como A es de rango máximo, R es regular, y además $R^{-1}\hat{\mathbf{b}}_{1,\ldots,n}$ se puede calcular por retrosustitución.

7.2. Factorización QR y ortogonalización de Gram-Schmidt

Consideramos factorizaciones de matrices cuadradas de la forma

$$A = QR$$

donde Q es ortogonal y R es triangular superior, y, más generalmente factorizaciones de matrices $m \times n$, $m \geq n$ de la forma

$$A = Q \underbrace{\begin{pmatrix} R \\ 0_{m-n,n} \end{pmatrix}}_{S}$$

donde Q es una matriz ortogonal $m \times m$ y R es triangular superior $n \times n$.

Estas factorizaciones siempre existen. Si notamos por $\mathbf{a}_k$ las columnas de A y por $\mathbf{q}_k$ las de Q tenemos

$$\begin{pmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{pmatrix} = \begin{pmatrix} \mathbf{q}_1 & \dots & \mathbf{q}_n \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ & r_{22} & \dots & r_{2n} \\ & & \vdots & \\ & & & r_{nn} \end{pmatrix},$$

es decir (ver fórmulas 1.2 y 1.1)

$$\begin{aligned} \mathbf{a}_1 &= r_{11}\mathbf{q}_1 \\ \mathbf{a}_1 &= r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 \\ &\dots \\ \mathbf{a}_n &= r_{1n}\mathbf{q}_1 + \dots + r_{n-1,n}\mathbf{q}_{n-1} + r_{n-1,n}\mathbf{q}_n. \end{aligned}$$

Observamos que

$$\mathbf{a}_k = \underbrace{r_{1k}\mathbf{q}_1 + \dots + r_{k-1,k}\mathbf{q}_{k-1}}_{P_{\langle \mathbf{q}_1,\dots,\mathbf{q}_{k-1}\rangle}\mathbf{a}_k} + \underbrace{r_{kk}\mathbf{q}_k}_{P_{\langle \mathbf{q}_1,\dots,\mathbf{q}_{k-1}\rangle^\perp}\mathbf{a}_k}.$$

Por tanto podemos obtener los vectores $\mathbf{q}_k$ y los coeficientes r_{kl} a partir de los vectores $\mathbf{a}_k$ mediante *ortogonalización de Gram-Schimidt*, que se corresponde con el siguiente algoritmo, en el que en la etapa k calculamos la k-ésima columna de Q y de R.

- $\mathbf{q}_1 = \dfrac{\mathbf{a}_1}{\|\mathbf{a}_1\|}$, $r_{11} = \|\mathbf{a}_1\|$.

- Para $k = 2, \ldots, n$
  - Para $j = 1, \ldots, k-1$, $r_{jk} = \mathbf{q}_j^* \mathbf{a}_k$.
  - $\tilde{\mathbf{q}}_k = \mathbf{a}_k - \sum_{j=1}^{k-1} r_{jk}\mathbf{q}_j$.
  - $\mathbf{q}_k = \dfrac{\tilde{\mathbf{q}}_k}{\|\tilde{\mathbf{q}}_k\|}$, $r_{kk} = \|\tilde{\mathbf{q}}_k\|$.

Si en alguna iteración del algoritmo el nuevo vector $\mathbf{a}_k$ pertenece al subespacio generado por los $\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}$ ($\tilde{\mathbf{q}}_k = \mathbf{0}$), basta tomar un $\mathbf{q}_k$ cualquiera ortogonal a los $\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}$ y $r_{kk} = 0$ para continuar el algoritmo. Si, por el contrario, las columnas de A son linealmente independientes, obtendremos una matriz R con elementos reales positivos en la diagonal principal.

El algoritmo anterior es numéricamente inestable (sensible a los efectos derivados de la representación finita de los números reales), por lo que resulta preferible el *algoritmo de Gram-Schmidt modificado*, en el que cada vez que se obtiene un nuevo vector $\mathbf{q}_k$ se resta a todos los vectores $\mathbf{a}_k$ que quedan por procesar su proyección ortogonal sobre el subespacio generado por $\mathbf{q}_k$. De esta forma calculamos en cada iteración k una columna de Q y una fila de R. El algoritmo, para matrices cuadradas regulares, es el siguiente.

- Inicializamos $\mathbf{a}_k^{(0)} = \mathbf{a}_k$, $k = 1, \ldots, n$.
- Para $k = 1, \ldots, n$,
  - $\mathbf{q}_k = \dfrac{\mathbf{a}_k^{(k-1)}}{\|\mathbf{a}_k^{(k-1)}\|}$, $r_{kk} = \|\mathbf{a}_k^{(k-1)}\|$.
  - Para $j = k+1, \ldots, n$,
    - $r_{kj} = \mathbf{q}_k^* \mathbf{a}_j^{(k-1)}$.
    - $\mathbf{a}_j^{(k)} = \mathbf{a}_j^{(k-1)} - r_{kj}\mathbf{q}_k$.

El coste computacional de ambos algoritmos es aproximadamente $\frac{2}{3}mn^2$ flops (cada *flop (floating point operation)* equivale a una suma y un producto en coma flotante) [10, p. 60].

Si $m > n$ hay que añadir columnas a la matriz de la izquierda hasta formar una matriz cuadrada ortogonal. Para ello basta tomar vectores aleatorios y calcular las proyecciones de cada uno de ellos sobre el complemento ortogonal del espacio generado por las columnas ya generadas.

La factorización QR de una matriz de rango máximo no es única salvo que exijamos que R tenga elementos positivos en su diagonal principal. De hecho, es fácil comprobar que

en cualquier factorización de matrices podemos multiplicar la columna k de la primera matriz por una constante z_k y la fila k de la segunda matriz por z_k^{-1} y obtenemos otra factorización. Dado que en el el caso de la factorización QR las columnas de la primera matriz son unitarias, los z_k deben ser de módulo unidad. Se puede demostrar que los coeficientes z_k unitarios son el único elemento que puede diferenciar dos factorizaciones QR de una matriz de rango máximo.

Ejercicios

7.1. Demostrar que si $A = QR = \tilde{Q}\tilde{R}$ son dos factorizaciones QR de la matriz de rango máximo A, con $Q = \begin{pmatrix}\mathbf{q}_1 & \dots & \mathbf{q}_n\end{pmatrix}$, $\tilde{Q} = \begin{pmatrix}\tilde{\mathbf{q}}_1 & \dots & \tilde{\mathbf{q}}_n\end{pmatrix}$, $R = \begin{pmatrix}\mathbf{r}_1^\top \\ \dots \\ \mathbf{r}_n^\top\end{pmatrix}$, $R = \begin{pmatrix}\tilde{\mathbf{r}}_1^\top \\ \dots \\ \tilde{\mathbf{r}}_n^\top\end{pmatrix}$, entonces existen constantes de módulo unidad z_k tales que $\tilde{\mathbf{q}}_k = z_k\mathbf{q}_k$ y $\tilde{\mathbf{r}}_k = z_k^{-1}\mathbf{r}_k$. Indicación: Operar por inducción, demostrándolo primero para $k = 1$ y luego, suponiéndolo cierto para k, demostrándolo para $k + 1$.

7.2. Implementar el algoritmo de Gram-Schmidt modificado para matrices cuadradas regulares.

7.3. Reflexiones de Householder

Una *reflexión respecto de un hiperplano* es una transformación lineal que deja invariantes los vectores del hiperplano y que transforma cada vector ortogonal al hiperplano en su opuesto.

Las reflexiones de Householder son reflexiones respecto de un hiperplano que transforman un vector dado $\mathbf{x}$ en un vector proporcional al vector de la base canónica $\mathbf{e}_1$ (figura 7.1).

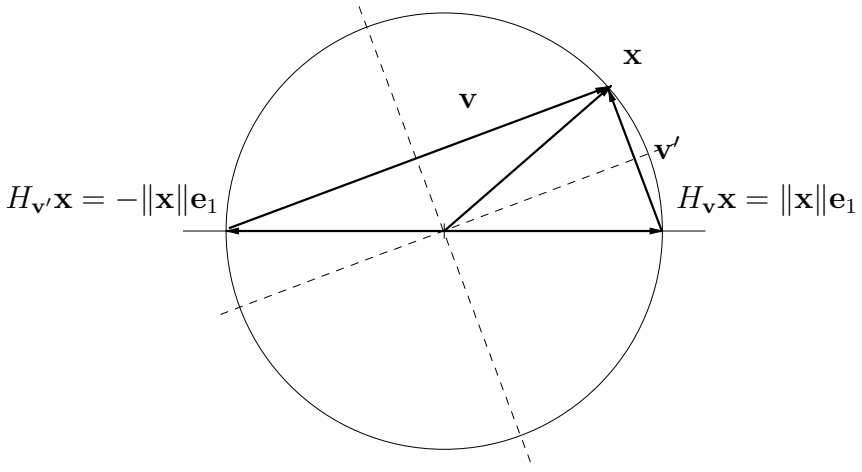

Figura 7.1: Reflexión de Householder.

La expresión de la matriz $H_{\mathbf{v}}$ de la reflexión respecto del hiperplano ortogonal al vector $\mathbf{v}$ es

$$H_{\mathbf{v}} = I - 2\frac{\mathbf{v}\mathbf{v}^*}{\mathbf{v}^*\mathbf{v}}, \tag{7.1}$$

y la reflexión que lleva $\mathbf{x}$ a $\operatorname{sgn}(x_1)\|\mathbf{x}\|\mathbf{e}_1$ corresponde al hiperplano perpendicular al vector

$$\mathbf{v} = \mathbf{x} + \operatorname{sgn}(x_1)\|\mathbf{x}\|\mathbf{e}_1. \tag{7.2}$$

De las dos posibilidades (la que transforma $\mathbf{x}$ en $\mathbf{e}_1$ y la que lo lleva a $-\mathbf{e}_1$) es preferible la indicada en la fórmula, porque evita la posibilidad de que $\mathbf{v}$ sea la diferencia entre dos vectores muy cercanos y por tanto su cálculo presente mucho error relativo.

Al ser isometrías, están dadas por matrices ortogonales. Permiten realizar la factorización QR de una matriz a base de ir transformando los vectores columna en vectores con ceros por debajo de la diagonal principal. Para una matriz 3×3:

$$\begin{pmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{pmatrix} \to \begin{pmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{pmatrix} \to \begin{pmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{pmatrix}.$$

En cada paso se utiliza una matriz ortogonal que aplica una reflexión de Householder al subvector columna que queda de la diagonal principal hacia abajo:

$$\begin{pmatrix} R_k & A \\ 0 & B \end{pmatrix} \to \begin{pmatrix} I & 0^\top \\ 0 & H_{\mathbf{v}} \end{pmatrix} \begin{pmatrix} R_k & A \\ 0 & B \end{pmatrix} = \begin{pmatrix} R_k & A \\ 0 & H_{\mathbf{v}}B \end{pmatrix}.$$

La triangularización de Householder tiene un coste computacional aproximado de $2mn^2 - \frac{2}{3}n^3$ flops [10, p. 75].

Ejercicios

7.3. Demostrar las fórmulas (7.1) y (7.2).

7.4. Implementar el algoritmo de factorización QR mediante reflexiones de Householder para matrices cuadradas regulares.

Capítulo 8

Otras factorización de matrices

8.1. Factorización LU

8.1.1. Resolución de sistemas de ecuaciones lineales mediante factorización LU

El conocido algoritmo de eliminación de Gauss para resolver sistemas de ecuaciones lineales

$$A\mathbf{x} = \mathbf{b}$$

da lugar a la factorización de la matriz inicial de la forma

$$A = LU$$

donde L *(lower)* en una matriz triangular inferior con unos en la diagonal principal y U *(upper)* es una matriz triangular superior.

Una vez obtenida la factorización, la solución del sistema

$$LU\mathbf{x} = \mathbf{b} \tag{8.1}$$

se realiza en dos etapas, obteniéndose primero $\mathbf{y} = U\mathbf{x}$ y luego $\mathbf{x}$. Al tratarse de matrices triangulares, basta aplicar dos veces retrosustitución.

La factorización LU de una matriz regular $A = (a_{ij})_{i,j=1,\dots,n}$ existe si y sólo si todas las submatrices $((a_{ij})_{i,j=1,\dots,m}$, $m < n$, (submatrices principales) son regulares [3, p. 158].

Ejercicios

8.1. Calcular aproximadamente el número de operaciones necesario para resolver el sistema (8.1).

8.1.2. Factorización LU básica

En cada etapa del algoritmo se utiliza una fila de la matriz para sumarsela, multiplicada por coeficientes adecuados, a las demás filas para ir formando columnas de ceros por debajo de la diagonal principal.

Lo vemos con un ejemplo (de [7, p. 279]) (repasar primero las fórmulas (1.3) y (1.4)).

$$\begin{aligned} A &= \begin{pmatrix} 2 & 4 & -5 \\ 6 & 8 & -1 \\ 4 & -8 & -3 \end{pmatrix} \\ B_{12} &= \begin{pmatrix} 2 & 4 & -5 \\ 0 & -4 & 16 \\ 4 & -8 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} A \\ B_{13} &= \begin{pmatrix} 2 & 4 & -5 \\ 0 & -4 & 16 \\ 0 & -16 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} B_{12} \\ B_{23} &= \begin{pmatrix} 2 & 4 & -5 \\ 0 & -4 & 16 \\ 0 & 0 & -57 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{pmatrix} B_{13} \end{aligned} \tag{8.2}$$

Recapitulando, tenemos

$$U = B_{23} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{pmatrix}}_{M_2} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{M_1 = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}} A. \tag{8.3}$$

donde hemos utilizado que el producto las matrices que crean ceros en una columna se obtiene simplemente acumulando los coeficientes, lo cual es consecuencia directa de la interpretación del producto de matrices como la actuación de la matriz de la izquierda sobre las filas de la de la derecha (fórmula (1.4)). De la misma forma se razona que la inversa de estas matrices se obtiene simplemente cambiando de signo los coeficientes fuera de la diagonal principal, lo que lleva a

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{pmatrix} U.$$

El producto de este tipo de matrices se calcula simplemente acumulando los elementos por debajo de la diagonal principal, puesto que

$$\begin{pmatrix} I_1 & \\ A & I_2 \end{pmatrix} \begin{pmatrix} I_1 & \\ & B \end{pmatrix} = \begin{pmatrix} I_1 & \\ A & B \end{pmatrix}. \tag{8.4}$$

Aplicado a nuestro ejemplo,

$$A = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 4 & 1 \end{pmatrix}}_{L} U.$$

Ejercicios

8.2. Calcular de forma aproximada el número de operaciones de la factorización LU de una matriz $n \times n$.

8.1.3. Factorización LU con pivoteo

Si el primer coeficiente no nulo de la fila que vamos a utilizar para crear una columna de ceros debajo de ella es muy pequeño, puede que el coeficiente con el que vamos a multiplicar la fila sea muy grande, y esto lleve, en el cálculo subsiguiente, a sumar números muy grandes con número mucho más pequeños. Y cuando utilizamos una representación con un número limitado de decimales esto puede llevar a que al sumar los dos números el resultado sea igual al más grande. En este apartado seguimos [7, p. 280] (ver también [7, problema 5.1-4]).

Para minimizar estos errores "de redondeo" resulta conveniente utilizar como columna "anuladora" la que presente mayor coeficiente. Esto se implementa conceptualmente premultiplicando la matriz por una *matriz de permutación* adecuada. Estas matrices corresponden a permutaciones de las filas de la matriz identidad y tienen el efecto de permutar las filas de la matriz a la que premultiplican. En nuestro caso las permutaciones son *transposiciones* (intercambios de dos elementos) y su misión es intercambiar la fila k ("anuladora") con la fila n, $n \geq k$, con coeficiente (n, k) de mayor valor absoluto.

Si en la factorización de una matriz 4×4 tuviéramos que utilizar transposiciones tendríamos en lugar de de (8.3)

$$U = M_3 P_3 M_2 P_2 M_1 P_1 A,$$

donde las P_i son las matrices de permutación (transposiciones) y las M_i son las matrices que van produciendo ceros en las columnas deseadas.

El pivoteo impide la consecución de una factorización LU, pero sigue permitiendo una factorización de la forma $PA = LU$, donde P es una matriz de permutación, que se obtiene como sigue. Como las transposiciones verifican $P_i^{-1} = P_i$,

$$\begin{aligned} U &= M_3P_3M_2P_2M_1P_1A = M_3(P_3M_2P_3)P_3P_2M_1P_1A \\ &= \underbrace{M_3}_{M_3'}\underbrace{(P_3M_2P_3)}_{M_2'}\underbrace{(P_3P_2M_1P_2P_3)}_{M_1'}\underbrace{(P_3P_2P_1)}_{P}A \\ PA &= \underbrace{M_1'^{-1}M_2'^{-1}M_3'^{-1}}_{L}U. \end{aligned}$$

L es triangular inferior como consecuencia de que las matrices M_i' mantienen la estructura de las M_i **Ejercicios**

(ejercicio 8.3).

Ejercicios

8.3. Realizar la factorización $PA = LU$ de la matriz A de (8.2). Explicar por qué el producto las matrices M_i' sigue siendo la estructura de las M_i, lo que permite aplicar (8.4).

8.2. Factorización de Cholesky

La factorización LU de una matriz hermítica definida positiva es un caso particular importante, que se requiere por ejemplo al resolver un problema de mínimos cuadrados mediante las ecuaciones normales. También aparecen en la solución numérica de ecuaciones en derivadas parciales. Podemos aprovechar la simetría de la matriz para reducir el coste del algoritmo y de paso obtener una factorización simétrica con mejores propiedades numéricas. Esta es la *factorización de Cholesky*. En esta sección seguimos [10, cap. 23].

La factorización de Cholesky de una matriz hermítica definida positiva $A = (a_{ij})$, $n \times n$, es de la forma

$$A = LL^* \tag{8.5}$$

donde L es triangular inferior con entradas no negativas en la diagonal principal.

Supongamos primero que $a_{11} = 1$. Aplicando eliminación gaussiana a la primera columna de A nos queda

$$\underbrace{\begin{pmatrix} 1 & \\ -\mathbf{w} & I \end{pmatrix}}_{B_1}\underbrace{\begin{pmatrix} 1 & \mathbf{w}^* \\ \mathbf{w} & K \end{pmatrix}}_{A} = \begin{pmatrix} 1 & \mathbf{w}^* \\ & K - \mathbf{w}\mathbf{w}^* \end{pmatrix}.$$

Visto que B_1 tiene la virtud de poner los ceros que queremos en la primera columna de la matrix, probamos a multiplicar A por la derecha por B_1^* con la esperanza de que ponga ceros de forma simétrica en primera fila. Y, efectivamente,

$$B_1 A B_1^* = \begin{pmatrix} 1 & \\ -\mathbf{w} & I \end{pmatrix} \begin{pmatrix} 1 & \mathbf{w}^* \\ \mathbf{w} & K \end{pmatrix} \begin{pmatrix} 1 & -\mathbf{w}^* \\ & I \end{pmatrix} = \begin{pmatrix} 1 & \\ & K - \mathbf{w}\mathbf{w}^* \end{pmatrix},$$

es decir,

$$A = \begin{pmatrix} 1 & \mathbf{w}^* \\ \mathbf{w} & K \end{pmatrix} = B_1^{-1} A B^{-*} = \begin{pmatrix} 1 & \\ \mathbf{w} & I \end{pmatrix} \begin{pmatrix} 1 & \\ & K - \mathbf{w}\mathbf{w}^* \end{pmatrix} \begin{pmatrix} 1 & \mathbf{w}^* \\ & I \end{pmatrix}.$$

Si $a_{11} \neq 1$, con un pequeño ajuste tenemos, definiendo $\alpha = \sqrt{a_{11}}$ (raíz cuadrada positiva (no puede ser $a_{11} < 0$ o no real, pues A no sería definida positiva)), tenemos

$$A = \begin{pmatrix} a_{11} & \mathbf{w}^* \\ \mathbf{w} & K \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha & \\ \mathbf{w}/\alpha & I \end{pmatrix}}_{H} \underbrace{\begin{pmatrix} 1 & \\ & K - \mathbf{w}\mathbf{w}^*/a_{11} \end{pmatrix}}_{C} \underbrace{\begin{pmatrix} \alpha & \mathbf{w}^*/\alpha \\ & I \end{pmatrix}}_{H^*}.$$

La matriz $A_1 = K - \mathbf{w}\mathbf{w}^*/a_{11}$ es definida positiva por serlo C, que es igual a $H^{-1}A(H^{-1})^*$. Si $n = 2$, ya hemos terminado, pues A_1 es un número real positivo y la factorización (8.5) es inmediata. Si $n > 2$, aplicamos la misma técnica a la matrix A_1, $(n-1) \times (n-1)$, y así sucesivamente hasta que obtenemos la factorización deseada.

La implementación se puede realizar almacenando solamente la parte triangular inferior de la matriz y aplicando a esta parte eliminación gaussiana. Con ello se consigue reducir aproximadamente a la mitad el número de operaciones de la factorización LU, con lo que el coste computacional de la factorización de Cholesky es $\sim \frac{1}{3}m^3$.

La factorización de Cholesky con elementos positivos en la diagonal principal de L es única.

La factorización de Cholesky también puede resultar ventajosa en términos de coste computacional, en relación a la factorización QR, en la resolución de problemas de mínimos cuadrados cuando la matriz A^*A es mucho más pequeña que la matriz A.

Capítulo 9

Cálculo de autovectores y autovalores

9.1. Introducción

El cálculo de autovalores de matrices genéricas no se puede hacer mediante algoritmos que terminen en un número prefijado de pasos porque ello sería contradictorio con la imposibilidad de resolver en radicales las ecuaciones generales de grado mayor o igual a cinco [10, p. 191].

Dado que los autovalores de una matriz aparecen en la diagonal principal de la matriz triangular de su factorización de Schur ($A = QTQ^*$, Q unitaria, T triangular superior), la obtención de esta factorización es también equivalente al cálculo de los autovalores de la matriz.

Sin embargo cualquier matriz se puede factorizar en un número finito de pasos como

$$A = QHQ^*$$

donde Q es unitaria y H es *Hessenberg superior*, es decir, tiene ceros por debajo de la diagonal por debajo de la diagonal principal. Para ello se van aplicando por la izquierda reflexiones de Householder que pongan ceros en las subcolumnas deseadas. Al aplicarlas también por la derecha estos ceros se respetan porque las reflexiones no afectan a las primeras filas igual que al aplicarlas por la izquierda no afectan a las primeras columnas (lo que no ocurriría si las reflexiones de Householder fueran las que van convirtiendo la matriz en triangular superior).

Si aplicamos este proceso a una matriz hermítica, H será hermítica además de Hessenberg superior, luego será tridiagonal.

Algunos algoritmos de cálculo de autovalores comienzan por factorizar de esta forma la matriz y luego aplican procesos iterativos para conseguir una matriz triangular superior similar a H.

9.2. Localización de autovalores

El teorama de los círculos de Gershgorin acota regiones del plano complejo en las que se encuentran los autovalores de una matriz. Concretamente, dada $A = (a_{ij})$, consideramos las componentes conexas de la unión de los círculos C_i de centro a_{ii} y radio $\sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|$ *(círculos de Gershgorin)*. El teorema establece que si una componente conexa es unión de m círculos, entonces contiene m autovalores contados con sus multiplicidades (con lo que el conjunto de los autovalores está contenido en la unión de todos los círculos).

Para demostrarlo veamos primero que el conjunto de los autovalores está contenido en la unión de los círculos. Si el autovector $\mathbf{x} = (x_1, \ldots, x_n)$ tiene autovalor λ y x_k es el coeficiente de $\mathbf{x}$ mayor valor absoluto, la acotación se deducr de la componente k de la ecuación $A\mathbf{x} = \lambda\mathbf{x}$:

$$\sum_{j=1}^{n} a_{kj}x_j = \lambda x_k \Leftrightarrow \sum_{\substack{j=1\\ j\neq k}}^{n} a_{kj}x_j = (\lambda - a_{kk})x_k$$

$$\Rightarrow |(\lambda - a_{kk})x_k| = |\lambda - a_{kk}|\,|x_k| \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{kj}|\,|x_j|$$

$$\Rightarrow |\lambda - a_{kk}| \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{kj}|\frac{|x_j|}{|x_k|} \leq \sum_{\substack{j=1\\ j\neq k}}^{n} |a_{kj}|.$$

Para demostrar la afirmación fuerte del teorema consideramos la curva de matrices $B(t) = D + t(A - D)$, $t \in [0, 1]$, donde $D = (d_{ij})$ es la matriz diagonal dada por $d_{ii} = a_{ii}$. Haremos uso de la continuidad de los autovalores con los coeficientes de la matriz (consecuencia de la continuidad de las raíces de los polinomios respecto de sus coeficientes). Obsérvese que $B(0)$ es la diagonal de A, $B(1) = A$ y los círculos de Gershgorin $C_i(t)$ de $B(t)$ tienen centro a_{ii} y radio tr_i, donde r_i es el radio del círculo para A.

El teorema es cierto para la matriz $B(0)$, puesto que los autovalores son los elementos de la diagonal de A y los círculos, de radio cero, son estos valores, y cada componente conexa corresponderá a un valor a_{ii}, con tantas componentes conexas como valores distintos haya en la diagonal de A. Para la matriz $B(1) = A$ sabemos que los autovalores están en la unión de los círculos.

Ahora dividimos la unión de los círculos de Gershgorin de $B(t)$ en dos componentes disjuntas, definiendo $K(t) = \cup_{k=1}^{m} C_{i_k}(t)$, $K' = \cup_{k=1}^{m'} C_{i'_k}(t)$. Cada una será unión de ciertas componentes conexas y lo que queremos demostrar es equivalente a que $K(t)$ contiene m autovalores y $K'(t)$ contiene m'. Ahora nos fijamos en una curva continua

$\lambda_k(t)$ dada por un autovalor de $B(t)$, con $\lambda_k(0) = a_{kk}$, $k \in \{i'_1, \ldots, i'_{m'}\}$ y llamamos $d(t)$ a la distancia entre $\lambda_k(t)$ y $K(t)$. Supongamos que, contrariamente a lo que queremos demostrar, $\lambda_i(1) \in K(1)$. Entonces $d(0) = d_0 > 0$ y $d(1) = 0$. Y observemos que $d(K(1), K'(1)) = d_1 > 0$ y que $d(K(t), K'(t))$ es no creciente, con lo que $d_0 \geq d_1 > 0$. Como $d(t)$ es continua y toma los valores d_0 y 0, en algún instante t_0 deberá tomar el valor intermedio $d_1/2$. Pero entonces para t_0 no está ni en $K(t_0)$ ni en $K'(t_0)$, lo que sabemos que no puede ocurrir. Por tanto hemos demostrado que las curvas como $\lambda_k(t)$ permanecen en la componente $K(t)$ o $K'(t)$ que les ha tocado inicialmente, de lo que se concluye el resultado buscado.

9.3. Iteración en las potencias y algoritmos relacionados

9.3.1. Iteración en las potencias

Esta técnica [10, p. 204] permite obtener el mayor autovalor (en valor absoluto) de una matriz hermítica A, suponiendo que sea simple, mediante la siguiente iteración, que parte de un vector unitario aleatorio $\mathbf{v}^{(0)}$, que no debe ser ortogonal al autovector asociado al autovalor buscado.

$$\mathbf{w} = A\mathbf{v}^{(k-1)}, \tag{9.1}$$

$$\mathbf{v}^{(k)} = \mathbf{w}/\|\mathbf{w}\|, \tag{9.2}$$

$$\lambda^{(k)} = \mathbf{v}^{(k)*} A \mathbf{v}^{(k)}. \tag{9.3}$$

La secuencia de los $\lambda^{(k)}$ converge al autovalor buscado. Tomando una base ortonormal de autovectores, $\mathbf{q}_i$, partimos del vector

$$\mathbf{v}^{(0)} = a_1\mathbf{q}_1 + \ldots + a_n\mathbf{q}_n$$

en el que suponemos $a_1 \neq 0$. Es fácil comprobar que

$$\mathbf{v}^{(k)} = \frac{A^k\mathbf{v}^{(0)}}{\|A^k\mathbf{v}^{(0)}\|} = \frac{a_1\lambda_1^k\mathbf{q}_1 + \ldots + a_n\lambda_n^k\mathbf{q}_n}{\sqrt{|a_1|^2\lambda_1^{2k} + \ldots + |a_n|^2\lambda_n^{2k}}},$$

y de aquí se obtiene (ejercicio)

$$\begin{aligned} |\lambda^{(k)} - \lambda_1| &\leq C\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}, \\ \|\mathbf{v}^{(k)} - (\pm\mathbf{q}_1)\| &\leq C'\left(\frac{\lambda_2}{\lambda_1}\right)^{k}. \end{aligned} \tag{9.4}$$

Ejercicios

9.1. Demostrar la acotación (9.4) indicando un valor explícito para la constante C.

9.3.2. Iteración inversa

Sabemos que si μ no está entre los autovalores λ_k de A, la matriz $(A-\mu I)^{-1}$ tiene los mismos autovectores que A y autovalores $1/(\lambda_k-\mu)$. Si conocemos un valor μ que esté más cerca de λ_K que de cualquier otro autovalor, la cantidad $1/(\lambda_K-\mu)$ será mucho más grande que las otras $1/(\lambda_k-\mu)$. Por tanto, aplicando iteración en las potencias a $(A-\mu I)^{-1}$ conseguiremos una convergencia

$$|\lambda^{(k)}-\lambda_1| \leq C\left|\frac{\mu-\lambda_J}{\mu-\lambda_K}\right|^{2k}$$

donde λ_J es el siguiente autovalor más cercano a μ.

El algoritmo modificado sustituye (9.1) por la resolución de

$$(A-\mu I)\mathbf{w} = \mathbf{v}^{(k-1)}.$$

9.3.3. Iteración en el cociente de Rayleigh

Este algoritmo deriva del anterior si sustituimos el parámetro fijo μ por la estimación $\lambda^{(k)}$ que disponemos del autovalor. Se puede demostrar que excepto para un conjunto de vectores iniciales de probabilidad cero (si los elegimos uniformemente en la esfera unidad) el algoritmo converge a un autovalor y, si el vector inicial está suficientemente cerca del autovector al que converge, la convergencia es cúbica [10, p. 206].

9.3.4. Iteración en las potencias para matrices no hermíticas

Los algoritmos anteriores también son válidos para matrices no hermíticas, pero el orden de convergencia cuadrático se convierte en lineal. Veamos cómo sería en este caso el algoritmo de iteración en las potencias. Supongamos que la matriz A tiene una base de autovectores $\mathbf{u}_j$, con autovalores λ_j, $|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|$. Entonces el vector de partida se puede escribir como

$$\mathbf{x}_0 = \sum_{j=1}^{n} a_j \mathbf{u}_j.$$

La iteración va a consistir

$$\mathbf{x}_{k+1} = A\mathbf{x}_k, \quad \lambda^{(k+1)} = \frac{x_{m,k+1}}{x_{m,k}},$$

donde $x_{m,k}$ es la componente m del vector $\mathbf{x}_k$ en la base canónica. Esta componente se elige arbitrariamente, pero es aconsejable tomar la componente de mayor valor absoluto después de varias iteraciones. Claramente

$$\mathbf{x}_k = \sum_{j=1}^{n} a_j \lambda_j^k \mathbf{u}_j = \lambda_1^k \left[a_1 \mathbf{u}_1 + \sum_{j=2}^{n} \left(\frac{\lambda_j}{\lambda_1} \right)^k a_j \mathbf{u}_j \right]$$

luego

$$\lambda^{(k+1)} = \lambda_1 \frac{a_1 u_{m,1} + \sum_{j=2}^{n} \left(\frac{\lambda_j}{\lambda_1} \right)^{k+1} a_j u_{m,j}}{a_1 u_{m,1} + \sum_{j=2}^{n} \left(\frac{\lambda_j}{\lambda_1} \right)^{k} a_j u_{m,j}}$$

y de ahi

$$|\lambda^{(k+1)} - \lambda_1| \leq C \left| \frac{\lambda_2}{\lambda_1} \right|^k$$

(ver [4]).

9.4. Algoritmo QR

Aunque sólo consideraremos el caso hermítico, el algoritmo QR proporciona los autovalores de la matriz si es regular, diagonalizable y la matriz de sus autovectores tiene factorización LU sin pivoteo (es decir, si sus menores principales son no nulos) [8, p. 378].

En las factorizaciones QR tomamos R con diagonal real positiva, de forma que esta factorización es única.

Algoritmo QR

- Inicializamos $A_1 = A$.
- Para $k = 1, 2, \ldots$
  - Calculamos la factorización $A_k = Q_k R_k$.
  - Calculamos el producto $A_{k+1} = R_k Q_k$.

Si A, con factorización de Schur $A = QTQ*$, verifica las condiciones enunciadas anteriormente, la secuencia A_k converge a T y la secuencia de los productos $Q_1Q_2\cdots Q_k$ converge a Q. Por tanto, si A es hermítica, estas secuencias de matrices convergen, respectivamente, a una matriz diagonal dada por los autovalores de A y a la matriz de los autovectores correspondientes.

Veamos la motivación intuitiva de esa afirmación para el caso hermítico. La demostración general se encuentra, por ejemplo, en [8, p. 378].

Supondremos además que los autovalores de A verifican $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_n|$. Consideremos la factorización QR de A^k para k elevado. Escribimos

$$A^k = A^k I = \begin{pmatrix} A^k\mathbf{e}_1 & \cdots & A^k\mathbf{e}_n \end{pmatrix}$$

y desarrollamos los vectores de la base canónica en términos de la base dada por los autovectores $\mathbf{q}_1, \ldots, \mathbf{q}_n$ de A:

$$\mathbf{e}_i = \alpha_{i1}\mathbf{q}_1 + \ldots + \alpha_{in}\mathbf{q}_n,$$

con lo que

$$A^k\mathbf{e}_i = \alpha_{i1}\lambda_1^k\mathbf{q}_1 + \ldots + \alpha_{in}\lambda_n^k\mathbf{q}_n.$$

Para k suficientemente grande podremos aproximar las columnas de A^k por

$$\begin{aligned} A^k\mathbf{e}_1 &\approx \alpha_{11}\lambda_1^k\mathbf{q}_1 \\ A^k\mathbf{e}_2 &\approx \alpha_{21}\lambda_1^k\mathbf{q}_1 + \alpha_{21}\lambda_1^k\mathbf{q}_1 \\ &\cdots \\ A^k\mathbf{e}_n &= \alpha_{n1}\lambda_1^k\mathbf{q}_1 + \ldots + \alpha_{nn}\lambda_n^k\mathbf{q}_n. \end{aligned}$$

con lo que su ortogonalización de Gram-Schmidt dará los autovectores de A, y por tanto para k suficientemente grande

$$A^k = Q^{(k)}R^{(k)} \Rightarrow Q^{(k)} \approx \begin{pmatrix} \mathbf{q}_1 & \cdots \mathbf{q}_n \end{pmatrix}.$$

Observemos ahora que nuestro algoritmo nos proporciona las factorizaciones QR de las potencias de A. En efecto, tenemos

$$\begin{aligned} A = A_1 &= Q_1R_1 \\ Q_2R_2 = A_2 &= R_1Q_1 = Q_1^*AQ_1 \\ Q_3R_3 = A_3 &= R_2Q_2 = Q_2^*A_2Q_2 = Q_2^*Q_1^*AQ_1Q_2 \\ &\cdots \\ Q_kR_k = A_k &= R_{k-1}Q_{k-1} = Q_k^*\cdots Q_1^*AQ_1\cdots Q_k \end{aligned}$$

mientras que

$$\begin{aligned}
A &= Q_1 R_1 \\
A^2 &= Q_1(R_1 Q_1)R_1 = Q_1 Q_2 R_2 R_1 \\
A^3 &= Q_1(R_1 Q_1)Q_2 R_2 R_1 = Q_1 Q_2(R_2 Q_2)R_2 R_1 = Q_1 Q_2 Q_3 R_3 R_2 R_1 \\
&\cdots \\
A^k &= Q_1 Q_2 \cdots Q_k R_k R_{k-1} \cdots R_1 \\
&\cdots
\end{aligned}$$

es decir,

$$Q^{(k)} = Q_1 Q_2 \cdots Q_k, \quad R^{(k)} = R_k R_{k-1} \cdots R_1.$$

y $A_k = Q^{(k)*} A Q^{(k)}$ se aproximará a la diagonalización de A según $Q^{(k)}$ tiende a la matriz de sus autovectores.

9.5. Coste computacional

El coste computacional de cada iteración de los algoritmos de iteración en las potencias es $O(n^2)$ (multiplicación matriz-vector).

El de cada iteración inversa es $O(n^3)$ (resolución de un sistema lineal), pero se reduce a $O(n^2)$ si la matriz se factoriza como QR o LU.

Si A es simétrica y se convierte antes en tridiagonal mediante transformaciones de Householder, cada iteración pasa a tener un coste $O(n)$.

Por lo que respecta al algoritmo QR, recordamos 10.4 que tanto el producto de dos matrices como la factorización QR tienen un coste $O(n^2)$.

La convergencia del algoritmo QR es lineal con constante $\max_j \frac{|\lambda_{j+1}|}{|\lambda_j|}$ [10, p. 218], pero esta constante se mejora con la versión modificada del algoritmo conocida como *algoritmo QR con desplazamientos* [10, p. 219], que integra en el esquema la idea de la iteración inversa (sección 9.3.2).

Ejercicios

9.2. Implementar cualquiera de los algoritmos de este capítulo.

Capítulo 10

Estabilidad de algoritmos numéricos

10.1. Modelo de aritmética de coma flotante

Los cálculos en un ordenador convencional que opera en coma flotante se puede modelar por las siguiente suposiciones, en las que interviene un parámetro característico ϵ_m *(ϵ de la máquina)*:

- Cada número x se representa mediante un valor
$$\mathrm{fl}(x) = x(1+\epsilon),\ |\epsilon| < \epsilon_m.$$
- El resultado de cada operación aritmética $*$ (suma, resta, multiplicación, división) entre dos números x e y que el ordenador representa de forma exacta ($x = \mathrm{fl}(x), y = \mathrm{fl}(y)$) verifica
$$x *_{\mathrm{fl}} y = (x * y)(1+\epsilon),\ |\epsilon| < \epsilon_m.$$

10.2. Estabilidad y retroestabilidad

Se dice que un algoritmo $\tilde{f}$ que implementa la función f es *estable* si existen C y C' tales que
$$\frac{\|\tilde{x} - x\|}{\|x\|} \leq C\epsilon_m \Rightarrow \frac{\|\tilde{f}(x) - f(\tilde{x})\|}{\|f(\tilde{x})\|} \leq C'\epsilon_m.$$

Se dice que un algoritmo $\tilde{f}$ que implementa la función f es *retroestable (backward stable)* si, para cada vector de datos x, $\tilde{f}(x)$ es tal que existe un vector $\tilde{x}$ cercano a x en el sentido de que para cierto C fijo
$$\frac{\|\tilde{x} - x\|}{\|x\|} \leq C\epsilon_m,$$

tal que

$$\tilde{f}(x) = f(\tilde{x}).$$

Por tanto un algoritmo retroestable nos da la solución exacta a un problema ligeramente perturbado respecto del que queremos resolver.

Por ejemplo, si $x = A$ es una matriz regular y $\tilde{f}(x) = (\tilde{Q}, \tilde{R})$ son las matrices de su factorización QR, si el algoritmo es retroestable obtenemos matrices tales que

$$\tilde{Q}\tilde{R} = A + \Delta A,\ \|\Delta A\| = \|\tilde{Q}\tilde{R} - A\| \leq C\epsilon_m\|A\|.$$

Si $x = A$ es la matriz regular del sistema $A\mathbf{y} = \mathbf{b}$, $\mathbf{y} = f(A)$ y $\tilde{\mathbf{y}} = \tilde{f}(A)$, el resultado verificará

$$(A + \Delta A)\tilde{\mathbf{y}} = \mathbf{b} \Rightarrow \|A\tilde{\mathbf{y}} - \mathbf{b}\| \leq \|\Delta A\|\|\tilde{\mathbf{y}}\| \leq C\epsilon_m\|A\|\|\tilde{\mathbf{y}}\|.$$

En cuanto a la distancia entre el resultado obtenido y el exacto tenemos, usando (6.1), que

$$\frac{\|\tilde{f}(x) - f(x)\|}{\|f(x)\|} = \frac{\|f(\tilde{x}) - f(x)\|}{\|f(x)\|} \leq \mathrm{cond}_x f \frac{\|\tilde{x} - x\|}{\|x\|} \leq C\,\mathrm{cond}_x f\,\epsilon_m.$$

10.3. Algunos resultados sobre estabilidad de algoritmos

Estos resultados se encuentran enunciados, unos demostrados y otros no, en [10].

- La factorización QR mediante reflexiones de Householder es retroestable [10, p. 116].
- En cuanto a la resolución de sistemas de ecuaciones
  - La resolución de un sistema triangular de ecuaciones (retrosustitución) es retroestable [10, p. 121].
  - La resolución de un sistema de ecuaciones mediante factorización QR es retroestable [10, p. 117].
  - La factorización LU es inestable para algunas matrices (raras) [10, p. 165].
  - Tanto la factorización de Cholesky como la resolución de sistemas de ecuaciones basada en ella son retroestables.
- En cuanto a la resolución del problema de mínimos cuadrados:
  - Mediante triangularización de Householder es retroestable [10, p. 138].

- Mediante ortogonalización de Gram-Schmidt modificado no lo es, pero el algoritmo puede modificarse para que lo sea [10, p. 140]. Sin embargo, requiere un número algo mayor de operaciones.
- Mediante SVD es también retroestable [10, p. 142].
- Sin embargo, la resolución directa de las ecuaciones normales puede ser inestable [10, p. 141].

El algoritmo QR es retroestable [10, p. 223].

10.4. Resumen de algoritmos de factorización

El cuadro 10.1 resume las características de los algoritmos de factorización que hemos estudiado.

Factorización	Algoritmo	Operaciones	Estabilidad
$A = QR$, $A \in M^{m\times n}$, $Q \in O_m$, $R \in U^{m\times n}$	Gram-Schmidt mod.	$2mn^2$	No
	Householder	$2mn^2 - \frac{2}{3}n^3$	RE
$A = LU$, $A \in M^{n\times n}$ $L \in L_1^{n\times n}$, $U \in U_{n\times n}$	Elim. gaussiana	$\frac{2}{3}n^3$	No
$A = LL^*$, $A \in H_n^+$ $L \in L^{n\times n}$	Fact. Cholesky	$\frac{1}{3}n^3$	RE

Cuadro 10.1: Factorizaciones de matrices que se implementan mediante algoritmos con un número finito de operaciones. $M^{m\times n}$: matrices $m\times n$, O_n: matrices ortogonales $n\times n$, $U^{m\times n}$: matrices triangulares superiores $m\times n$, $L^{m\times n}$: matrices triangulares inferiores $m\times n$, $L_1^{m\times n}$: matrices de $L^{m\times n}$ con unos en la diagonal principal, H_n^+: matrices hermíticas (semi)definidas positivas, RE: retroestable.

Bibliografía

[1] G. Golub, C. Van Loan, *Matrix Computation*, John Hopkins Univ. Press, Londres, RU.

[2] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2^a ed., Cambridge Univ. Press, RU, 2003.

[3] R. A. Horn, C. R. Johnon, *Matrix Analysis*, Cambridge Univ. Press, EEUU, 1985.

[4] E. Isaacson, H. Bishop, *Analysis of Numerical Methods*, 1966, Dover, EEUU.

[5] K. V. Mardia, J. T. Kent, J. M. Bibby, "Multivariate Analysis", Academic Press, Londres 1979.

[6] L. Merino, E. Santos, *Algebra Lineal con métodos elementales,* Thomson Paraninfo, 1997.

[7] T. K. Moon, W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice-Hall, EEUU, 2000.

[8] R. Plato, *Concise Numerical Mathematics*, Graduate Texts in Mathematics, American Mathematical Society, 2003.

[9] W. Press et al., *Numerical Recipes in C*, Prentice-Hall, EEUU.

[10] L. N. Trefethen, D. Bau, III, *Numerical Linear Algebra*, SIAM, EEUU, 1997.